

Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Kätrin Suvi

**Mitmemõõtmeline asümmeetriline $t_{p,\nu}$ -jaotus ja
koopula**

Kindlustus- ja finantsmatemaatika

Magistritöö (30 EAP)

Juhendaja: vanemteadur Tõnu Kollo

Tartu 2019

Mitmemõõtmeline asümmeetriline $t_{p,\nu}$ -jaotus ja koopula

Magistritöö

Kätrin Suvi

Lühikokkuvõte. Magistritöös antakse ülevaade asümmeetriliste elliptilise jaotuste klassi kuuluvast asümmeetrilisest mitmemõõtmelisest t -jaotusest. Tuletatakse selle jaotuse kirjeldamiseks vajalikud χ -pöördjaotuse momendid ja t -jaotuse parameetrite hinnangud ning esitatakse selle jaotuse baasil konstrueeritud koopula. Töö praktilises osas rakendatakse asümmeetrilist t -koopulat liikluskindlustuskahjude modelleerimiseks.

CERCS teaduseriala: Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatematika (P160)

Märksõnad: Asümmeetriline normaaljaotus, asümmeetriline t -jaotus, koopula, kindlustuskahjude modelleerimine, mitmemõõtmelised jaotused.

Multivariate skew-normal $t_{p,\nu}$ -distribution and copula

Master's Thesis

Kätrin Suvi

Abstract. This Master's thesis gives an overview of the multivariate skew t -distribution, which belongs to the class of skew elliptical distributions. The moments of inverse χ -distribution are derived and presented in order to describe skew t -distribution. Estimates of the parameters of the skew t -distribution are found and copula based on this distribution is introduced. In the practical part, skew t -copula is applied for modelling motor insurance claims.

CERCS research specialisation: Statistics, operation research, programming, actuarial mathematics (P160)

Keywords: Copula, modelling of insurance claims, multivariate distributions, skew-normal distribution, skew t -distribution.

Sisukord

Sissejuhatus	4
1 Mitmemõõtmeline asümmeetriline $t_{p,\nu}$-jaotus	6
1.1 Pidevad elliptilised jaotused	6
1.2 Mitmemõõtmeliste jaotuste arvkarakteristikud	8
1.3 Asümmeetriline normaaljaotus $SN(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$	10
1.3.1 $SN(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ parameetrite hindamine	12
1.4 χ_ν - pöördjaotus	13
1.4.1 χ_ν -pöördjaotuse momendid	15
1.5 Asümmeetriline t - jaotus $St_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$	17
1.5.1 $St_{p,\nu}(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ parameetrite hindamine	20
1.5.2 Parameetrite hinnangute käitumine	21
2 Koopula mõiste	23
2.1 Peamised mõisted	24
2.2 Sklari teoreem	25
2.3 Asümmeetriline $t_{p,\nu}$ -koopula	26
2.4 Korrelatsioon	27
3 Kahjude analüüs	29
3.1 Andmestiku kirjeldus	29
3.2 Jaotuste sobitamine andmetele	31
3.2.1 Kahjusumma suurus	32
3.2.2 Kahjumenetluse aeg	35
3.3 Ühisjaotuse leidmine	37
3.3.1 Algoritm asümmeetrilise $t_{p,\nu}$ -koopula väärtuste simuleerimiseks	37
3.3.2 Ühisjaotusest simuleerimine	38
Kokkuvõte	41
Summary	42
Viited	43
Lisad	45
Lisa 1. Simuleerimisülesande R-i kood	45
Lisa 2. Kahjude analüüsi R-i kood	46

Sissejuhatus

Magistritöö eesmärgiks on saada ülevaade mitmemõõtmelise $t_{p,\nu}$ -jaotuste baasil moodustatud asümmeetrilisest mitmemõõtmelisest $t_{p,\nu}$ -koopulast ja rakendada seda kindlustusandmetel. Asümmeetriline $t_{p,\nu}$ -jaotus on oluline mudel finants- ja kindlustusmatemaatikas, kuna lisaks andmete ebasümmeetriale võimaldab arvesse võtta tunnuste sõltuvust väga suurte väärtuste korral, ehk niinimetatud sabasõltuvust. Eriti väärtuslikuks muudab asümmeetrilise $t_{p,\nu}$ -jaotuse asjaolu, et temast moodustatud ühisjaotuse - *koopula* - abil on võimalik moodustada mitmemõõtmeline jaotus, kus erinevat tüüpi marginaaljaotused on ühendatud mitmemõõtmeliseks jaotuseks nii, et arvesse on võetud ka tunnustevaheline sõltuvusstruktuur.

Ülesehituselt on antud töö järgnev. Esimene peatükk keskendub mitmemõõtmelise $t_{p,\nu}$ -jaotuse tutvustamisele. Esmalt on antud lühiülevaade elliptilistest jaotustest. Seejärel on defineeritud mitmemõõtmeline asümmeetriline normaaljaotus ning toodud välja avaldised jaotuse keskvaartusvektori ning kovariatsioonimaatriksi leidmiseks. Samuti on leitud punkthinnangud jaotuse parameetritele momentide meetodil ning esitatud graafilised näited jaotuse käitumisest kahemõõtmelisel juhul. Järgneb χ -pöördjaotuse tutvustus ning tihedusfunktsiooni integreerimise teel on leitud jaotuse keskvaartuse ning dispersiooni avaldised. Viimaks on tutvustatud asümmeetrilist $t_{p,\nu}$ -jaotust. Leitud on jaotuse keskvaartusvektori ning kovariatsioonimaatriksi avaldised kasutades kahe eelnevalt tutvustatud jaotuse omadusi. Lisaks on leitud punkthinnangud jaotuse parameetritele momentide meetodil.

Teises peatükis antakse lühiülevaade koopulateooria ajaloolisest arengust. Esitatakse koopula definitsioon ning olulisemad mõisted ja teoreemid. Viimases paragrahvis tutvustatakse asümmeetrilist $t_{p,\nu}$ -koopulat.

Kolmas peatükk on rakenduslik. On teostatud kahjude analüüs ning modelleeritud andmeid $t_{2,4}$ -koopula abil. Kasutatud on aastast 2001 pärinevaid Läti kindlustuseltsi liikluskindlustuse kahjude andmeid. Esmalt on sobitatud tunnuste empiirilistele jaotustele neli erinevat teoreetilist jaotust. Sobitamisel on kasutatud Gamma-, Pareto, Weibulli ning lognormaalset jaotust, hinnangud jaotuste parameetritele on leitud suurima tõepära meetodil. Sobivaima teoreetilise jaotuse valimisel on kasutatud Kolmogorov-Smirnovi kooskõlatesti teststatistiku väärtusi ning informatsiooni-kriteeriume *AIC* ja *BIC*. Seejärel on leitud kahe tunnuse ühisjaotus asümmeetrilise $t_{2,4}$ -koopula abil.

Magistritöö kirjutamiseks on kasutatud tekstitöötlus programmi *TexMaker*. Kõik

analüüsid, graafikud, empiirilistele andmetele jaotuste sobitamine ning ühisjaotuse leidmine on läbi viidud vabavaralise statistikapaketiga *RStudio*. Samuti on asümmeetrilise $t_{p,\nu}$ -koopula simuleerimise algoritm kirja pandud R-i funktsioonidena.

Käesolevaga tänab autor magistritöö juhendajat rohkete nõuannete ning suuniste, paranduste ja soovitude, eriliselt aga meeldiva koostöö ja pühendatud ajast eest.

1 Mitmemõõtmeline asümmeetriline $t_{p,\nu}$ -jaotus

Asümmeetrilised p -mõõtmelised t -jaotused kuuluvad asümmeetriliste elliptiliste jaotuste hulka. Need saadakse elliptilistest jaotustest teisenduste abil. Pidades silmas p -mõõtmelist t -jaotust vabadusastmete arvuga ν kasutame tähistust $t_{p,\nu}$. Mitmemõõtmelised $t_{p,\nu}$ -jaotused on olnud teadlaste huviorbiidis juba viimased 80 aastat [9]. Kõige enam on mitmemõõtmelised $t_{p,\nu}$ -jaotused kasutatud leidnud klassikalistes statistilistes modelleerimisülesannetes, kus jaotuste sabad on raskemad kui normaaljaotusel [9]. Samuti on rakendatud t -jaotust statistikute robustsuse uurimisel. Mitmemõõtmeline asümmeetriline $t_{p,\nu}$ -jaotus on huvi pakkunud just seetõttu, et pakub enam paindlikkust, võttes arvesse nii sabade raskust kui ka andmete paiknemise ebasümmeetrilisust. Paindlikkuse tõttu on antud jaotus hinnatud mitmetes praktilistes rakendustes, eriti finantsmatemaatikas.

Nii nagu mitmemõõtmelisel normaaljaotusel on mitmeid esitusviise, on ka asümmeetrilist $t_{p,\nu}$ -jaotust defineeritud mitmeti. Ülevaade erinevatest lähenemistest on esitatud monograafias [9], § 5. Antud magistritöös keskendume A. Azzalini ja A. Capitanio 2003. aastal artiklis [1] välja toodud lähenemisele, kus asümmeetriline $t_{p,\nu}$ -jaotus on defineeritud kahe juhusliku suuruse korrutisena, millest üks on asümmeetrilise normaaljaotusega ning teine χ -pöördjaotusega. Järgnevalt anname kõigepealt lühiülevaate elliptilistest pidevatest jaotustest, seejärel kirjeldame kaht sagedamini kasutatavat jaotust. Lõpuks anname ülevaate ka uuritavast mitmemõõtmelisest asümmeetrilisest t -jaotusest ning leiame hinnangud parameetritele kasutades asümmeetrilise normaaljaotuse ja χ -pöördjaotuse karakteristikuid.

1.1 Pidevad elliptilised jaotused

Järgnev paragrahv tugineb raamatule [8], § 2.3. Klassikaline mitmemõõtmeline analüüs on üles ehitatud vaatlusandmete normaaljaotuse eeldusele. Reaalsete andmete korral on see eeldus väga harva täidetud. Üheks mitmemõõtmelise analüüsi arendamise peamiseks eesmärgiks on olnud normaaljaotuse üldistamine ehk normaaljaotuse eeldusest vabanemine. Suurt tähelepanu on pöördunud elliptilistele jaotustele. Need jaotused on köitnud erinevate statistikute tähelepanu mitmel põhjusel. Esiteks just seetõttu, et antud klass sisaldab normaaljaotuseid. Teiseks on mitmed normaaljaotuse korral saadud tulemused ning järeldused kergesti üle kantavad elliptilistele mudelitele.

Elliptiliste jaotuste hulgas kannab *sfääriline* jaotus sama tähtsat rolli, nagu standardne mitmemõõtmeline normaaljaotus $N_p(\mathbf{0}, \mathbf{I})$ mitmemõõtmeliste normaaljaotuste $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ hulgas.

Definitsioon 1.1.1. *Öeldakse, et juhuslik p -vektor \mathbf{X} on **sfäärilise jaotusega**, kui \mathbf{X} ja $\boldsymbol{\Gamma}^T \mathbf{X}$ on sama jaotusega iga $p \times p$ ortogonaalse matriksi $\boldsymbol{\Gamma}$ korral.*

Kui \mathbf{X} on pideva sfäärilise jaotusega juhuslik vektor, siis võrduse $\boldsymbol{\Gamma}^T \mathbf{X} = \mathbf{I}_p$ tõttu sõltub vektori tihedusfunktsioon argumentist suuruse $\mathbf{x}^T \mathbf{x}$ kaudu. Levinuimad sfäärilised jaotused:

- mitmemõõtmeline normaaljaotus $N_p(\mathbf{0}, \mathbf{I}_p)$ tihedusfunktsiooniga

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2\sigma^2} \mathbf{x}^T \mathbf{x}};$$

- mitmemõõtmeline t_p -jaotus vabadusastmete arvuga ν ning tihedusfunktsiooniga

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{\frac{p}{2}}} \left(1 + \frac{1}{\nu} \mathbf{x}^T \mathbf{x}\right)^{-\frac{\nu+p}{2}}.$$

Definitsioon 1.1.2. *Öeldakse, et juhuslik vektor \mathbf{X} on **elliptilise jaotusega**, parameetritega $\boldsymbol{\mu} : p \times 1$ ning $\mathbf{V} : p \times p$, kui \mathbf{X} on sama jaotusega, kui*

$$\boldsymbol{\mu} + \mathbf{A}\mathbf{Y},$$

kus \mathbf{Y} on sfäärilise jaotusega ning $\mathbf{A} : p \times k$, $\mathbf{A}\mathbf{A}^T = \mathbf{V}$. Matriksi \mathbf{V} astak on k , $r(\mathbf{V}) = k$.

Märgime $\mathbf{X} \sim E_p(\boldsymbol{\mu}, \mathbf{V})$. Pidevate mittesingulaarsete elliptiliste jaotuste korral on $\mathbf{A} : p \times p$ täisastakuga p .

Näited kõige enam kasutatavatest elliptilistest jaotustest:

- mitmemõõtmeline normaaljaotus $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ kuulub elliptiliste jaotuste hulka, sest kui $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, siis vektor \mathbf{X} on esitatav kujul $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$, kus $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ ja $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$ ning $\mathbf{A} : p \times p$ on pööratav;
- mitmemõõtmelise t -jaotuse parameetritega $\boldsymbol{\mu}$ ja $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$, vabadusastmete arvuga ν , saame sarnase teisenduse abil: $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$, kus \mathbf{Y} on sfäärilise jaotusega, tihedusfunktsiooniga (1.5.1) ja $\mathbf{A} : p \times p$ on pööratav. Sel juhul kirjutame $\mathbf{X} \sim t_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Olgu meil elliptilise jaotusega vektor $\mathbf{X} \sim E_p(\boldsymbol{\mu}, \mathbf{V})$. Selleks, et tihedus eksisteeriks, on vajalik, et $r(\mathbf{V}) = p$.

Sel juhul peab definitsioonis esinev maatriks \mathbf{A} olema täisastakuga $p \times p$ ruutmaatriks. Vektori \mathbf{X} tihedusfunktsioon $f_{\mathbf{X}}(\mathbf{X})$ avaldub järgmiselt:

$$f_{\mathbf{X}}(\mathbf{x}) = c_p |\mathbf{V}|^{\frac{1}{2}} g((\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})),$$

kus c_p on normeeriv konstant ja funktsioon $g(\cdot)$ rahuldab tingimust

$$\int_0^\infty y^{\frac{p}{2}-1} g(y) dy < \infty.$$

Käesolevas ülevaates on esitatud vaid paar näidet elliptiliste jaotuste hulgast. Tegelikult on neid jaotuseid rohkem. Kui selles paragrahvis tegime põgusalt tutvust mitmemõõtmelise normaaljaotusega, siis järgnevalt vaatame mitmemõõtmeliste jaotuste arvkarakteristikuid ning seejärel asümmeetrilist mitmemõõtmelist normaaljaotust.

1.2 Mitmemõõtmeliste jaotuste arvkarakteristikud

Nii nagu ühemõõtmeliste juhuslike suuruste korral iseloomustavad ka mitmemõõtmelisi juhuslikke vektoreid üldkogumi arvkarakteristikud. Järgnev ülevaade on refereeritud mitmemõõtmelise statistika veebikonspektist [11]. Sarnaselt ühemõõtmelise juhusliku suuruse keskväärtusele ning dispersioonile on mitmemõõtmelisel juhul peamisteks karakteristikuteks paiknevuse ning hajuvuse arvkarakteristikud. Juhusliku p -vektori \mathbf{Y} keskväärtusvektor $E\mathbf{Y}$ on vektori komponentide keskväärtuste vektor kujul

$$E\mathbf{Y} = \begin{pmatrix} EY_1 \\ EY_2 \\ \dots \\ EY_p \end{pmatrix}.$$

Vektori \mathbf{Y} keskväärtusvektor leidub vaid juhul, kui iga $i \in \{1, 2, \dots, p\}$ korral keskväärtus EY_i leidub. Juhusliku vektori dispersioonimaatriks $D\mathbf{Y}$, mis on sarnane

ühemõõtmelise jaotuse dispersioonile, on kujul

$$D\mathbf{Y} = \begin{pmatrix} DY_1 & cov(Y_1, Y_2) & \dots & cov(Y_1, Y_p) \\ cov(Y_2, Y_1) & DY_2 & \dots & cov(Y_2, Y_p) \\ \dots & \dots & \dots & \dots \\ cov(Y_p, Y_1) & cov(Y_p, Y_2) & \dots & DY_p \end{pmatrix}.$$

Nii nagu ka keskvärtusvektori korral, leidub dispersioonimaatriks vaid juhul, kui eksisteerivad kõik kovariatsioonid $cov(Y_i, Y_j)$ ning dispersioonid DY_i , $i, j \in \{1, 2, \dots, p\}$. Dispersioonimaatriksi $D\mathbf{Y}$ peadiagonaalil on juhusliku vektori komponentide dispersioonid $DY_i = E[Y_i - EY_i]^2$, väljaspool diagonaali juhusliku vektori komponentide vahelised kovariatsioonid $cov(Y_i, Y_j) = E[(Y_i - EY_i)(Y_j - EY_j)]$. Dispersioonimaatriksi saame esitada ka keskvärtusena teatavast juhuslikust maatriksist

$$D\mathbf{Y} = E[(\mathbf{Y} - E\mathbf{Y})(\mathbf{Y} - E\mathbf{Y})^T]. \quad (1.2.1)$$

Dispersioonimaatriks on sümmeetriline ning positiivselt määratud $p \times p$ -maatriks. Tähistame dispersioonimaatriksi Σ .

Mitmemõõtmelises statistikas lähtutakse sageli järelduste tegemisel tunnuste vahelistest seosesetest. Lisaks keskvärtusvektorile ning dispersioonimaatriksile on tähtis informatsiooniallikas ka korrelatsioonimaatriks \mathbf{R} , mis on kujul

$$\mathbf{R} = \begin{pmatrix} 1 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & 1 & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p1} & \sigma_{p2} & \dots & 1 \end{pmatrix}, \quad (1.2.2)$$

kus σ_{ij} on tunnuste Y_i ja Y_j vaheline korrelatsioonikordaja,

$$\sigma_{ij} = \frac{cov(Y_i, Y_j)}{\sqrt{DY_i DY_j}}.$$

Märgime, et kovariatsioonimaatriksi Σ kaudu saame esitada korrelatsioonimaatriksi \mathbf{R} järgnevalt:

$$\mathbf{R} = \Sigma_d^{-\frac{1}{2}} \Sigma \Sigma_d^{-\frac{1}{2}}, \quad (1.2.3)$$

kus Σ_d märgib diagonaliseeritud maatriksit Σ :

$$\Sigma_d = \begin{pmatrix} DY_1 & 0 & \dots & 0 \\ 0 & DY_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & DY_p \end{pmatrix}.$$

Siin ja edaspidi eeldame, et diagonaliseerimine rakendatakse maatriksile enne teisi maatriksoperatsioone. Korrelatsioonimaatriks on sümmeetriline positiivselt määratud maatriks, tema peadiagonaalil asuvad alati ühed.

1.3 Asümmeetriline normaaljaotus $SN(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$

Asümmeetrilise normaaljaotuse defineerime järgides artiklit [6].

Definitsioon 1.3.1. *Olgu meil juhuslik p -mõõtmeline vektor $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)^T$. Siis vektor \mathbf{Y} on **p -mõõtmelise asümmeetrilise normaaljaotusega** parameetritega $\boldsymbol{\mu}$, Σ ja $\boldsymbol{\alpha}$, kui tema tihedusfunktsioon avaldub kujul*

$$f_{\mathbf{Y}}(\mathbf{y}) = 2f_{N_p(\boldsymbol{\mu}, \Sigma)}(\mathbf{y})\Phi(\boldsymbol{\alpha}^T(\mathbf{y} - \boldsymbol{\mu})), \quad (1.3.1)$$

kus $f_{N(\boldsymbol{\mu}, \Sigma)}(\mathbf{y})$ on p -mõõtmelise normaaljaotuse $N_p(\boldsymbol{\mu}, \Sigma)$ tihedusfunktsioon ning $\Phi(\cdot)$ on standardse normaaljaotuse $N(0, 1)$ jaotusfunktsioon.

Tähistame tihedusfunktsiooniga (1.3.1) vektori \mathbf{Y} jaotuse vastavalt $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$, kus $\boldsymbol{\mu} \in \mathbb{R}^p$ on asukoha parameeter, $\boldsymbol{\alpha} \in \mathbb{R}^p$ on asümmeetrilisust kirjeldav parameeter ning $\Sigma : p \times p$ on positiivselt määratud hajuvusparameeter.

Märgime ära, et vektor $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$ on esitatav ka kujul [7]

$$\mathbf{Y} = \begin{cases} \mathbf{Z}, & \alpha^T(\mathbf{Z} - \boldsymbol{\mu}) > Z_0, \\ -\mathbf{Z}, & \text{muul juhul,} \end{cases} \quad (1.3.2)$$

kus $\mathbf{Z} \sim N_p(\boldsymbol{\mu}, \Sigma)$ ja $Z_0 \sim N(0, 1)$.

Asümmeetrilise normaaljaotuse $SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ momente genereeriv funktsioon avaldub kujul ([2], § 5.1.2):

$$M(\mathbf{t}) = 2e^{t^T \boldsymbol{\mu} + \frac{1}{2} t^T \boldsymbol{\Sigma} t} \Phi \left(\frac{\boldsymbol{\alpha}^T \boldsymbol{\Sigma} \mathbf{t}}{\sqrt{1 + \boldsymbol{\alpha}^T \boldsymbol{\Sigma} \boldsymbol{\alpha}}} \right).$$

Jaotuse esimesed momendid saame leida, kui võtame maatrikstuletised momente genereerivast funktsioonist. Asümmeetrilise normaaljaotuse keskvärtusvektori ja kovariatsioonimaatriksi tähistame vastavalt $E\mathbf{Y}$ ja $D\mathbf{Y}$. Nende tuletuskäik on esitatud artiklis [6]:

$$\begin{aligned} E\mathbf{Y} &= \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} \boldsymbol{\delta}, \\ D\mathbf{Y} &= \boldsymbol{\Sigma} - \frac{2}{\pi} \boldsymbol{\delta} \boldsymbol{\delta}^T, \end{aligned}$$

kus $\boldsymbol{\delta}$ on kujul:

$$\boldsymbol{\delta} = \frac{\boldsymbol{\Sigma} \boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}^T \boldsymbol{\Sigma} \boldsymbol{\alpha}}}. \quad (1.3.3)$$

Viimasest avaldisest saame algebraliste teisenduste kaudu avaldada $\boldsymbol{\alpha}$ järgmiselt:

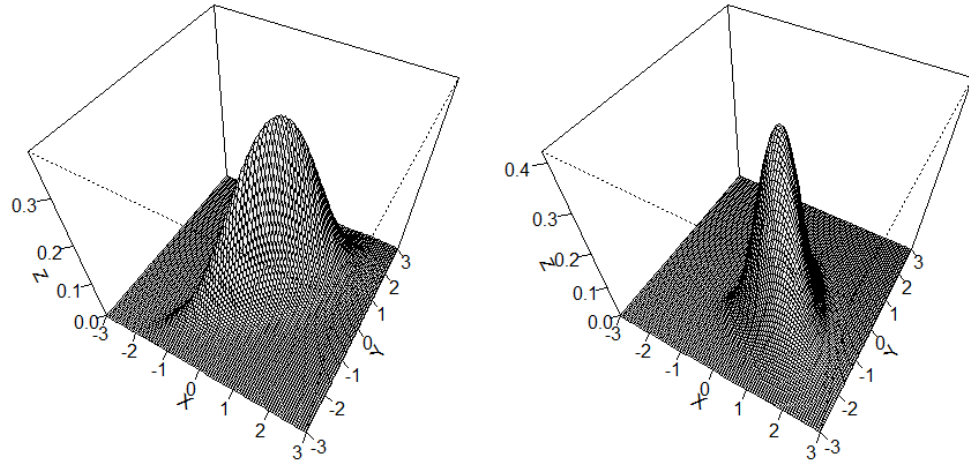
$$\boldsymbol{\alpha} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}}{\sqrt{1 - \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}}}. \quad (1.3.4)$$

Edaspidi vaatame kaheparameetrilist asümmeetrilist normaaljaotust, kui $\boldsymbol{\mu} = 0$, tähistame $SN_p(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$. Sel juhul avalduvad jaotuse keskvärtusvektor ning dispersioonimaatriksi järgmiselt:

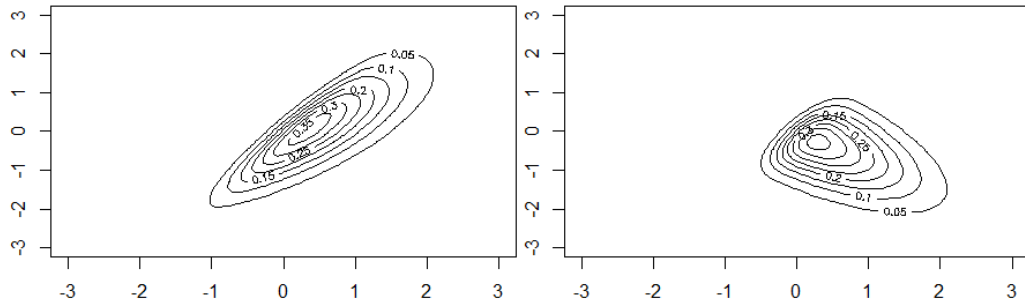
$$E\mathbf{Y} = \sqrt{\frac{2}{\pi}} \boldsymbol{\delta}, \quad (1.3.5)$$

$$D\mathbf{Y} = \boldsymbol{\Sigma} - \boldsymbol{\delta} \boldsymbol{\delta}^T \frac{2}{\pi}. \quad (1.3.6)$$

Mitmemõõtmelise SN_p -jaotuse kuju sõltub parameetrite $\boldsymbol{\Sigma}$ ning $\boldsymbol{\alpha}$ koostõjust. Järgnevad graafikud kirjeldavad jaotuse $SN_2(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ kuju erinevate parameetrite väärtuste korral. Joonise 1 vasakpoolsel graafikul on kujutatud jaotust korrelatsioonimaatriksiga $\mathbf{R}_1 = \begin{pmatrix} 1 & 0,7 \\ 0,7 & 1 \end{pmatrix}$ ning $\boldsymbol{\alpha} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$. Parempoolsel graafikul kujutatud jaotuse asümmeetriaparameter $\boldsymbol{\alpha}$ on sama kui vasakpoolsel, kuid $\mathbf{R}_2 = \begin{pmatrix} 1 & -0,7 \\ -0,7 & 1 \end{pmatrix}$. Joonisel 2 on paremaks ülevaateks esitatud ka nende jaotuste tihedusfunktsiooni samatõenäosusjooned.



Joonis 1: Kahe asümmeetrilise normaaljaotuse tihedusfunktsiooni võrdlus.



Joonis 2: Jaotuste tihedusfunktsioonide samatõenäosusjooned.

1.3.1 $SN(\Sigma, \alpha)$ parameetrite hindamine

Järgnevalt leiame $SN_p(\Sigma, \alpha)$ -jaotuse parameetrite hinnangud. Asendades $E\mathbf{Y}$ ja $D\mathbf{Y}$ nihketa hinnangutega $\bar{\mathbf{Y}}$ ja $\mathbf{S}_\mathbf{Y}$ saame momentide meetodi rakendamisel võr-
dused:

$$\widehat{E\mathbf{Y}} = \bar{\mathbf{Y}} = \sqrt{\frac{2}{\pi}} \hat{\boldsymbol{\delta}}, \quad (1.3.7)$$

$$\widehat{D\mathbf{Y}} = \mathbf{S}_\mathbf{Y} = \hat{\Sigma}_\mathbf{Y} - \bar{\mathbf{Y}}\bar{\mathbf{Y}}^T, \quad (1.3.8)$$

kust saame leida hinnangud parameetritele Σ ja α ning vektorile δ vastavalt:

$$\begin{aligned}\hat{\delta} &= \bar{\mathbf{Y}} \sqrt{\frac{\pi}{2}}, \\ \hat{\Sigma}_{\mathbf{Y}} &= \mathbf{S}_{\mathbf{Y}} + \bar{\mathbf{Y}} \bar{\mathbf{Y}}^T, \\ \hat{\alpha} &= \frac{\hat{\Sigma}_{\mathbf{Y}}^{-1} \hat{\delta}}{\sqrt{1 - \hat{\delta}^T \hat{\Sigma}_{\mathbf{Y}}^{-1} \hat{\delta}}} = \frac{[\mathbf{S}_{\mathbf{Y}} + \bar{\mathbf{Y}} \bar{\mathbf{Y}}^T]^{-1} \bar{\mathbf{Y}} \sqrt{\frac{\pi}{2}}}{\sqrt{1 - [\bar{\mathbf{Y}} \sqrt{\frac{\pi}{2}}]^T [\mathbf{S}_{\mathbf{Y}} + \bar{\mathbf{Y}} \bar{\mathbf{Y}}^T]^{-1} [\bar{\mathbf{Y}} \sqrt{\frac{\pi}{2}}]}} \\ &= \frac{[\mathbf{S}_{\mathbf{Y}} + \bar{\mathbf{Y}} \bar{\mathbf{Y}}^T]^{-1} \bar{\mathbf{Y}} \sqrt{\frac{\pi}{2}}}{\sqrt{1 - \frac{\pi}{2} \bar{\mathbf{Y}}^T [\mathbf{S}_{\mathbf{Y}} + \bar{\mathbf{Y}} \bar{\mathbf{Y}}^T]^{-1} \bar{\mathbf{Y}}}} = \frac{[\mathbf{S}_{\mathbf{Y}} + \bar{\mathbf{Y}} \bar{\mathbf{Y}}^T]^{-1} \bar{\mathbf{Y}}}{\sqrt{\frac{2}{\pi} - \bar{\mathbf{Y}}^T [\mathbf{S}_{\mathbf{Y}} + \bar{\mathbf{Y}} \bar{\mathbf{Y}}^T]^{-1} \bar{\mathbf{Y}}}}.\end{aligned}$$

1.4 χ_ν - pöördjaotus

Statistikas ning tõenäosusteoorias on χ_ν^2 -jaotus väga laialdaselt kasutatav jaotus, kus ν tähistab vabadusastmete arvu. Enim kasutatakse χ_ν^2 - jaotust hüpoteeside testimisel ja usalduspiiride leidmisel.

Definitsioon 1.4.1. Olgu X_1, X_2, \dots, X_ν sõltumatud standardse normaalajaotusega juhuslikud suurused, $X_i \sim N(0, 1)$. Siis $Y = \sum_{i=1}^\nu X_i^2$ on χ_ν^2 -**jaotusega**, vabadusastmete arvuga ν , tähistame $Y \sim \chi_\nu^2$.

χ_ν^2 - jaotuse tihedusfunktsioon on järgmine ([12], lk. 330):

$$f_\nu(x) = \begin{cases} x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}} \cdot \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma(\frac{\nu}{2})}, & \text{kui } x > 0; \\ 0, & \text{mujal,} \end{cases}$$

kus $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ on gammafunktsioon.

Olgu gammajaotus defineeritud kujul nagu antud töö paragrahvis 3.2. Siis χ_ν^2 -jaotus on Γ -jaotuse erijuht, kui $\alpha = \frac{\nu}{2}$ ja $\beta = 2$. Seega kui $\mathbf{Y} \sim \chi_\nu^2$, siis $\mathbf{Y} \sim \Gamma\left(\frac{\nu}{2}, 2\right)$.

Definitsioon 1.4.2. Olgu $Z^2 \sim \chi_\nu^2$. Siis ütleme, et Z on χ_ν -**jaotusega**, vabadusastmete arvuga ν ning tähistame $Z \sim \chi_\nu$.

χ_ν -jaotuse tihedusfunktsioon on järgmisel kujul ([12], lk. 333) :

$$f_\nu(x) = \begin{cases} x^{\nu-1} e^{-\frac{x^2}{2}} \cdot \frac{1}{2^{\frac{\nu}{2}-1} \Gamma(\frac{\nu}{2})}, & \text{kui } x > 0; \\ 0, & \text{mujal.} \end{cases}$$

Definitsioon 1.4.3. Olgu $Z \sim \chi_\nu$. Siis ütleme, et $V_\nu = \frac{1}{Z}$ on χ_ν -**pöördjaotusega**, vabadusastmete arvuga ν ning tähistame $V_\nu \sim I\chi_\nu$.

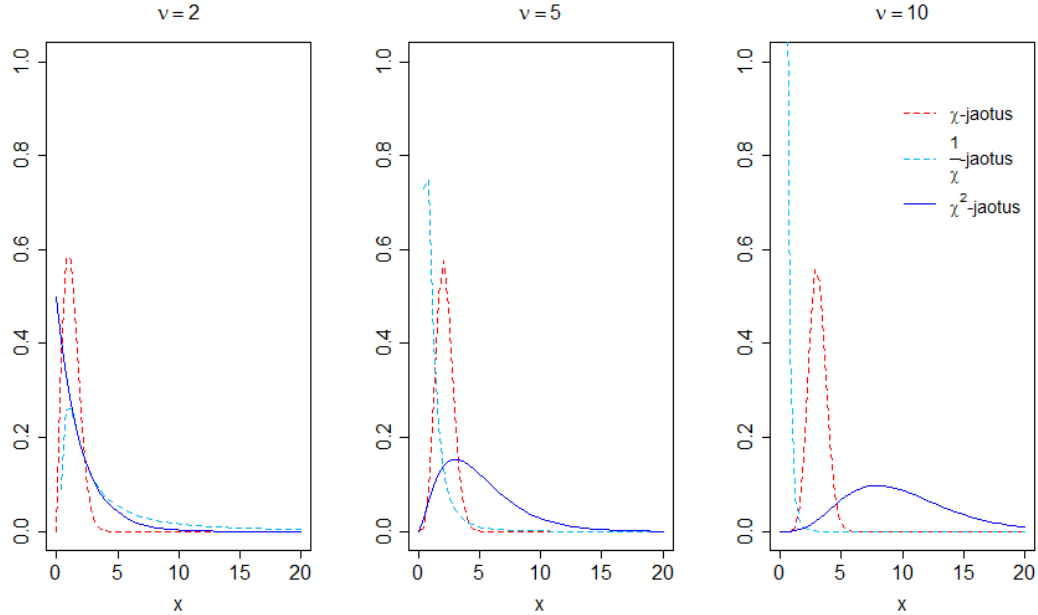
χ_ν -pöördjaotuse tihedusfunktsiooni leiame χ_ν -jaotuse tihedusfunktsioonist:

$$f_\nu(y) = f_{Z_\nu} \left(\frac{1}{x} \right) \cdot \left| \frac{dx}{dy} \right| = y^{-\nu+1} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu}{2}-1} \Gamma\left(\frac{\nu}{2}\right)} \cdot \left| \left(-\frac{1}{y^2} \right) \right|.$$

Saame

$$f_\nu(y) = y^{-\nu-1} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu}{2}-1} \Gamma\left(\frac{\nu}{2}\right)}. \quad (1.4.1)$$

Joonisel 3 on kolm graafikut, kus on välja toodud χ_ν^2 -jaotuse, χ_ν -jaotuse ja $I\chi_\nu$ -jaotuse tihedusfunktsioonid erinevate vabadusastmete arvu ν korral, vastavalt 2, 5, 10. Näeme, et χ_ν^2 -jaotuse korral muutub tihedusfunktsiooni kuju suurema vabadusastmete arvu korral lamedamaks ning sarnasemaks normaaljaotusega. χ_ν -jaotuse korral muutub vabadusastmete arvu suurenedes jaotusfunktsiooni saba kergemaks. $I\chi_\nu$ -jaotuse korral aga on tõenäosusmassi põhiosa nulli lähedale kogunenud.



Joonis 3: χ_ν^2 -jaotuse, χ_ν -jaotuse ja $I\chi_\nu$ -jaotuse võrdlus erinevate vabadusastmete arvu ν korral.

1.4.1 χ_ν -pöördjaotuse momendid

Järgnevalt leiame $I\chi_\nu$ -jaotuse keskväärtuse, dispersiooni ning k -ndat järku momendi tihedusfunktsiooni integreerimise teel.

Lause 1.4.1. χ_ν -pöördjaotuse momendid on:

$$EV_\nu = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)}, \text{ kui } \nu > 1;$$

$$EV_\nu^2 = \frac{1}{\nu-2}, \text{ kui } \nu > 2;$$

$$EV_\nu^k = \frac{\Gamma\left(\frac{\nu-k}{2}\right)}{2^{k/2}\Gamma\left(\frac{\nu}{2}\right)}, \text{ kui } \nu > k.$$

Tõestus. Olgu $V_\nu \sim I\chi_\nu$. Siis tema tihedusfunktsioon avaldub kujul (1.4.1).

Kirjutame välja ka $V_{\nu-1}$ tihedusfunktsiooni:

$$f_{V_{\nu-1}}(x) = x^{-(\nu-1)-1} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu-1}{2}-1}\Gamma\left(\frac{\nu-1}{2}\right)} = x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu-3}{2}}\Gamma\left(\frac{\nu-1}{2}\right)}$$

Jaotuse momendid saame leida integreerimise teel:

$$m_k(V_\nu) = EV_\nu^k = \int_0^\infty x^k f_{V_\nu}(x) dx.$$

Järgnevalt on tõestustes kasutatud tihedusfunktsiooni omadust

$$\int_0^\infty f_{V_\nu}(x) dx = 1.$$

Esmalt leiame keskväärtuse ehk esimest järku momendi:

$$\begin{aligned} m_1(V_\nu) = EV_\nu &= \int_0^\infty x \cdot x^{-\nu-1} e^{-\frac{1}{2x^2}} \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} dx = \int_0^\infty x^{-\nu} e^{-\frac{1}{2x^2}} \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} dx \\ &= \int_0^\infty x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu-3}{2}+\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)} dx \\ &= \int_0^\infty x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{1}{\sqrt{2} \cdot 2^{\frac{\nu-3}{2}}\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)} dx \\ &= \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu-3}{2}}\Gamma\left(\frac{\nu-1}{2}\right)} = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)}. \end{aligned}$$

Siit saame:

$$m_1(V_\nu) = EV_\nu = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)}. \quad (1.4.2)$$

Seejärel leiame teist järku momendi:

$$\begin{aligned} m_2(V_\nu) &= EV_\nu^2 = \int_0^\infty x^2 x^{-\nu-1} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} dx = \int_0^\infty x \cdot x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} dx \\ &= \int_0^\infty x \cdot x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2} \cdot 2^{\frac{\nu-3}{2}} \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{\nu-1}{2}\right)} dx \\ &= \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty x \cdot x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu-3}{2}} \Gamma\left(\frac{\nu-1}{2}\right)} dx \\ &= EV_\nu \int_0^\infty x \cdot f_{V_{\nu-1}} dx = EV_\nu \cdot EV_{\nu-1}. \end{aligned}$$

Näeme, et teist järku moment avaldub kahe esimest järku momendi korrutisena.

Kasutame eespool leitud EV_ν avaldisi:

$$m_2(V_\nu) = EV_\nu^2 = EV_\nu \cdot EV_{\nu-1} = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu-1}{2}\right)} = \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{2 \cdot \Gamma\left(\frac{\nu}{2}\right)}$$

Siinkohal kasutame gammafunktsiooni omadust $\Gamma(z+1) = z\Gamma(z)$, ehk antud juhul $\Gamma\left(\frac{\nu}{2}\right) = \frac{\nu-2}{2}\Gamma\left(\frac{\nu-2}{2}\right)$. Seega saame:

$$m_2(V_\nu) = EV_\nu^2 = \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{2 \cdot \Gamma\left(\frac{\nu}{2}\right)} = \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{2 \cdot \frac{\nu-2}{2}\Gamma\left(\frac{\nu-2}{2}\right)} = \frac{1}{\nu-2} \quad (1.4.3)$$

Leiame ka k -ndat järku momendi:

$$\begin{aligned} m_k(V_\nu) &= EV_\nu^k = \int_0^\infty x^k \cdot x^{-\nu-1} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} dx \\ &= \frac{2^{\frac{\nu-1}{2}-1}\Gamma\left(\frac{\nu-1}{2}\right)}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty x^{k-1} \cdot x^{-\nu} e^{-\frac{1}{2x^2}} \cdot \frac{1}{2^{\frac{\nu-1}{2}-1}\Gamma\left(\frac{\nu-1}{2}\right)} dx \\ &= \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty x^{k-1} \cdot f_{V_{\nu-1}}(x) dx = EV_\nu \cdot EV_{\nu-1}^{k-1}. \end{aligned}$$

Siit aga saame, et

$$\begin{aligned} EV_\nu^k &= EV_\nu EV_{\nu-1} EV_{\nu-2}^{\nu-2} = EV_\nu EV_{\nu-1} EV_{\nu-2}^{\nu-2} \cdot \dots \cdot EV_{\nu-(k-1)} \\ &= \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{2^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{2^{\frac{1}{2}}\Gamma\left(\frac{\nu-2}{2}\right)} \cdot \frac{\Gamma\left(\frac{\nu-3}{2}\right)}{2^{\frac{1}{2}}\Gamma\left(\frac{\nu-3}{2}\right)} \cdot \dots \cdot \frac{\Gamma\left(\frac{\nu-k}{2}\right)}{2^{\frac{1}{2}}\Gamma\left(\frac{\nu-(k+1)}{2}\right)} = \frac{\Gamma\left(\frac{\nu-k}{2}\right)}{2^{\frac{k}{2}}\Gamma\left(\frac{\nu}{2}\right)}. \end{aligned}$$

□

Järeldus 1.4.1. IX_ν dispersioon avaldub kujul:

$$DV_\nu = EV_\nu^2 - [EV_\nu]^2 = \frac{1}{\nu-2} - \left[\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)} \right]^2 = \frac{1}{\nu-2} - \frac{1}{2} \left[\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right]^2. \quad (1.4.4)$$

1.5 Asümmeetriline t - jaotus $St_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$

Mitmemõõtmelisel $t_{p,\nu}$ -jaotusel eksisteerib mitmeid erinevaid definitsioone ning üldistusi. Lähtume järgmisest mitmemõõtmelise $t_{p,\nu}$ -jaotuse definitsioonist ([9], §1.1):

Definitsioon 1.5.1. Olgu meil juhuslik p -vektor $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)^T$. Siis \mathbf{Y} on **p -mõõtmelise t -jaotusega** vabadusastmete arvuga ν , keskväärtsvektoriga $\boldsymbol{\mu}$ ning hajuvusparameetriga $\boldsymbol{\Sigma}$, kui tema tihedusfunktsioon avaldub kujul

$$t_{p,\nu}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{(\pi\nu)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left[1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{\nu} \right]^{-\frac{\nu+p}{2}}. \quad (1.5.1)$$

Tähistame $\mathbf{Y} \sim t_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Selle jaotuse korral on kovariatsioonimaatriks $\boldsymbol{\Sigma}$. Märkime, et kovariatsioonimaatriksist $\boldsymbol{\Sigma}$ saame leida korrelatsioonimaatriksi \mathbf{R} võrduse (1.2.3) kaudu.

Järgnevalt defineerime asümmeetrilise $t_{p,\nu}$ -jaotuse, järgides A. Azzalini ja A. Capitanio esitust raamatus [9], §5.9.

Definitsioon 1.5.2. Juhuslik p -vektor $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ on **p -mõõtmelise asümmeetrilise t -jaotusega**, vabadusastmete arvuga ν ning parameetritega $\boldsymbol{\mu}, \boldsymbol{\Sigma}$, ja $\boldsymbol{\alpha}$, kui tema tihedusfunktsioon avaldub kujul

$$g_{p,\nu}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) = 2t_{p,\nu}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) T_{1,p+\nu} \left[\boldsymbol{\alpha}^T (\mathbf{x} - \boldsymbol{\mu}) \sqrt{\frac{p+\nu}{\mathbf{Q} + \nu}}; p + \nu \right], \quad (1.5.2)$$

kus $\mathbf{Q} = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$, $T_{1,p+\nu}$ on vabadusastmete arvuga $p + \nu$ ühemõõtmelise t_ν -jaotuse jaotusfunktsioon, $t_{p,\nu}$ on p -mõõtmelise t -jaotuse tihedusfunktsioon

vabadusastmete arvuga ν .

Tähistame $\mathbf{X} \sim St_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$, kus $\boldsymbol{\alpha} \in \mathbb{R}^p$ on asümmeetrilisust määrav parameeter, $\boldsymbol{\mu} \in \mathbb{R}^p$ on asukoha parameeter, $\boldsymbol{\Sigma}$ on hajuvust kirjeldav positiivselt määratud $p \times p$ -maatriks. Antud töös käsitleme olukorda, kui $\boldsymbol{\mu} = 0$ ning seega tähistame $\mathbf{X} \sim St_\nu(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$.

Käesoleva magistritöö alguses sai märgitud, et antud töös kasutame asümmeetrilise $t_{p,\nu}$ -jaotuse definitsiooni, mille kohaselt asümmeetriline $t_{p,\nu}$ -jaotus on asümmeetrilise normaaljaotusega juhusliku vektori ning χ_ν -pöördjaotusega juhusliku suuruse korrutis.

Definitsioon 1.5.3. Olgu juhuslik vektor \mathbf{Y} p -mõõtmelise asümmeetrilise normaaljaotusega, $\mathbf{Y} \sim SN_p(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ ning juhuslik suurus Z χ_ν -jaotusega, $Z \sim \chi_\nu$, sõltumatu vektorist \mathbf{Y} . Siis

$$\mathbf{X} = \sqrt{\nu} \frac{\mathbf{Y}}{Z} \sim St_{p,\nu}(\boldsymbol{\Sigma}, \boldsymbol{\alpha}).$$

Tähistame juhusliku suuruse $V = \frac{1}{Z}$. Tänu juhusliku vektori \mathbf{Y} ja juhusliku suuruse V sõltumatusele saame asümmeetrilise $t_{p,\nu}$ -jaotusega vektori \mathbf{X} keskvaartusvektori $E\mathbf{X}$ avaldada juhusliku vektori \mathbf{Y} keskvaartusvektori ja juhusliku suuruse V keskvaartuse korrutisena:

$$E\mathbf{X} = E\left(\sqrt{\nu} \frac{\mathbf{Y}}{Z}\right) = E\left(\sqrt{\nu} \mathbf{Y} V\right) = \sqrt{\nu} E\mathbf{Y} E V.$$

Kasutades asümmeetrilise mitmemõõtmelise normaaljaotuse keskvaartusvektori (1.3.5) ning χ_ν - pöördjaotuse keskvaartuse (1.4.2) võrdusi saame:

$$E\mathbf{X} = \sqrt{\nu} \sqrt{\frac{2}{\pi}} \boldsymbol{\delta} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)} = \sqrt{\frac{\nu}{\pi}} \boldsymbol{\delta} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} = \sqrt{\frac{\nu}{\pi}} \frac{\boldsymbol{\Sigma} \boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}^T \boldsymbol{\Sigma} \boldsymbol{\alpha}}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}.$$

Dispersioonimaatriksi leiame kasutades paragrahvis 1.2 välja toodud valemit (1.2.1):

$$\begin{aligned} D\mathbf{X} &= E\left[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T\right] = E\left[\mathbf{X}\mathbf{X}^T - \mathbf{X}(E\mathbf{X}^T) - (E\mathbf{X})\mathbf{X}^T + (E\mathbf{X})(E\mathbf{X}^T)\right] \\ &= m_2(\mathbf{X}) - (E\mathbf{X})(E\mathbf{X}^T) = E\left(\sqrt{\nu} \mathbf{Y} V \sqrt{\nu} \mathbf{Y}^T V\right) - \sqrt{\nu} E V E \mathbf{Y} \sqrt{\nu} E V E \mathbf{Y}^T \\ &= \nu \left[E\left(V^2 \mathbf{Y} \mathbf{Y}^T\right) - (E V)^2 E \mathbf{Y} E \mathbf{Y}^T\right] = \nu \left[E V^2 E\left(\mathbf{Y} \mathbf{Y}^T\right) - (E V)^2 E \mathbf{Y} E \mathbf{Y}^T\right] \\ &= \nu \left[m_2(V) m_2(\mathbf{Y}) - (E V)^2 E \mathbf{Y} E \mathbf{Y}^T\right] \end{aligned}$$

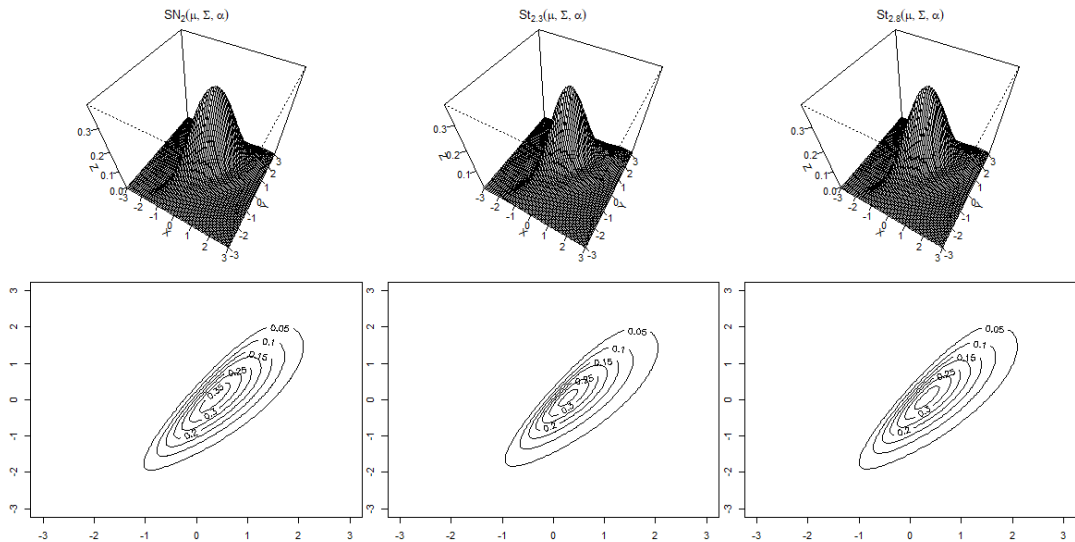
Kasutades teadmist, et $m_2(\mathbf{Y}) = \boldsymbol{\Sigma}$ ning eelnevalt leitud asümmeetrilise mitmemõõtmelise normaaljaotuse keskvaartusvektori (1.3.5) ning lauses 1.4.1 tõestatud

χ_ν - pöördjaotuse momente saame:

$$D\mathbf{X} = \nu \left[\frac{1}{\nu-2} \mathbf{\Sigma} - \left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right)^2 \frac{1}{\pi} \boldsymbol{\delta} \boldsymbol{\delta}^T \right],$$

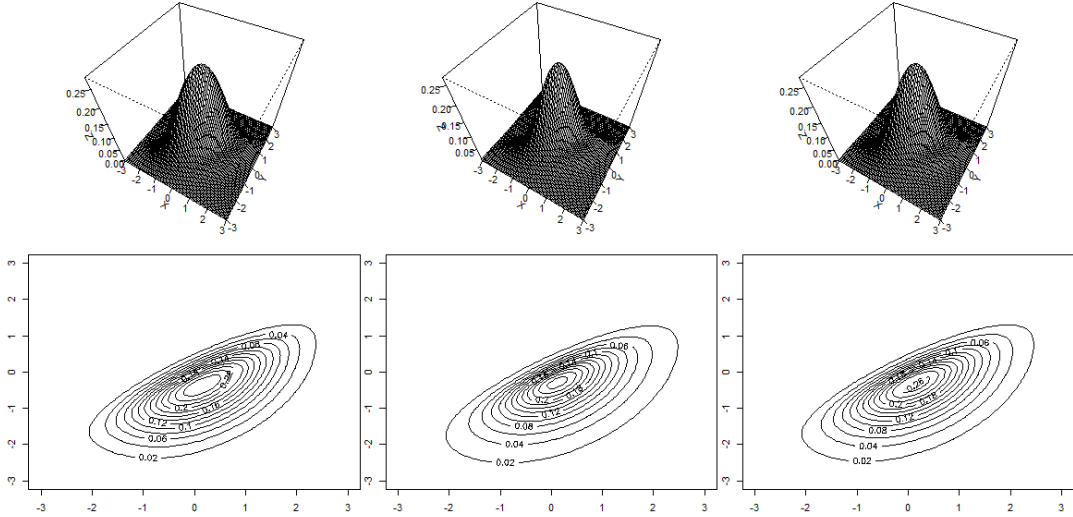
kus $\boldsymbol{\delta}$ on antud vördusega (1.3.3).

Järgnevatel joonistel 4 ja 5 on välja toodud $SN_2(\mathbf{R}, \boldsymbol{\alpha})$ -jaotuse ning $St_{2,\nu}(\mathbf{R}, \boldsymbol{\alpha})$ -jaotuste ruumilised tihedusfunktsiooni graafikud ning tihedusfunktsiooni samatõenäosusjooned. Joonisel 4 on parameetrite väärtused vastavalt $\boldsymbol{\alpha} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$ ning korrelatsioonimaatriks $\mathbf{R} = \begin{pmatrix} 1 & 0,7 \\ 0,7 & 1 \end{pmatrix}$. Vasakpoolsed graafikud on SN_2 -jaotuse kohta, keskmised joonised $St_{2,3}$ -jaotuse graafikud ning parempoolsed $St_{2,8}$ -jaotuse graafikud. Näeme, et mida suurem on vabadusastmete ν arv, seda sarnasem on asümmeetriline t -jaotus asümmeetrilisele normaaljaotusele.



Joonis 4: Jaotuste $SN_2(\mathbf{R}, \boldsymbol{\alpha})$, $St_{2,3}(\mathbf{R}, \boldsymbol{\alpha})$ ning $St_{2,8}(\mathbf{R}, \boldsymbol{\alpha})$ võrdlus parameetrite $\boldsymbol{\alpha} = (5, -3)^T$ ning $r_{11} = r_{22} = 1$, $r_{12} = r_{21} = 0,7$ korral.

Joonisel 5 on välja toodud samuti 3 erinevat jaotust nagu joonisel 4. Sel korral on parameetrite väärtused vastavalt $\boldsymbol{\alpha} = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$ ning korrelatsioonimaatriks $\mathbf{R} = \begin{pmatrix} 1 & 0,35 \\ 0,35 & 1 \end{pmatrix}$. Näeme, et graafikutel on hajuvus mõlemas suunas suurem, kuna $\boldsymbol{\alpha}$ on esimese tunnuse korral väiksem.



Joonis 5: Jaotuste $SN_2(\mathbf{R}, \boldsymbol{\alpha})$, $St_{2,3}(\mathbf{R}, \boldsymbol{\alpha})$ ning $St_{2,8}(\mathbf{R}, \boldsymbol{\alpha})$ võrdlus parameetrite $\boldsymbol{\alpha} = (2, -3)^T$ ning $r_{11} = r_{22} = 1$, $r_{12} = r_{21} = 0$, 35 korral.

1.5.1 $St_{p,\nu}(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ parameetrite hindamine

Asendades $E\mathbf{X}$, $D\mathbf{X}$ nihketa hinnangutega vastavalt $\overline{\mathbf{X}}$ ja $\mathbf{S}_{\mathbf{X}}$ ning kasutades töös eelnevalt leitud võrdusi (1.3.7), (1.3.8), (1.4.2), (1.4.3) saame momentide meetodi rakendamisel järgmised võrdused:

$$\widehat{E\mathbf{X}} = \overline{\mathbf{X}} = \frac{\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\alpha}}}{\sqrt{1 + \hat{\boldsymbol{\alpha}}^T \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\alpha}}}} \sqrt{\frac{\nu}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, \quad (1.5.3)$$

$$\widehat{D\mathbf{X}} = \mathbf{S}_{\mathbf{X}} = \nu \left[m_2(V)m_2(\mathbf{Y}) - (EV)^2 E\mathbf{Y}E\mathbf{Y}^T \right] = \nu \cdot \frac{1}{\nu-2} \hat{\boldsymbol{\Sigma}} - \overline{\mathbf{X}}\overline{\mathbf{X}}^T.$$

Viimasest võrdusest saame avaldada hinnangu parameetrile $\boldsymbol{\Sigma}$:

$$\hat{\boldsymbol{\Sigma}} = \frac{\nu-2}{\nu} \left(\mathbf{S}_{\mathbf{X}} + \overline{\mathbf{X}}\overline{\mathbf{X}}^T \right). \quad (1.5.4)$$

Hinnangu parameetrile $\boldsymbol{\alpha}$ saame leida algebraliste teisenduste kaudu võrdusest (1.5.3) asendades sisse ka $\hat{\boldsymbol{\Sigma}}$:

$$\begin{aligned}
\hat{\alpha} &= \frac{\hat{\Sigma}^{-1} \bar{\mathbf{X}} \sqrt{\frac{\pi}{\nu}} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})}}{\sqrt{1 - \left[\bar{\mathbf{X}} \sqrt{\frac{\pi}{\nu}} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})} \right]^T \hat{\Sigma}^{-1} \bar{\mathbf{X}} \sqrt{\frac{\pi}{\nu}} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})}}} \\
&= \frac{\frac{\nu}{\nu-2} (\mathbf{S}_{\mathbf{X}} + \bar{\mathbf{X}} \bar{\mathbf{X}}^T)^{-1} \bar{\mathbf{X}} \sqrt{\frac{\pi}{\nu}} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})}}{\sqrt{1 - \left[\bar{\mathbf{X}} \sqrt{\frac{\pi}{\nu}} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})} \right]^T \frac{\nu}{\nu-2} (\mathbf{S}_{\mathbf{X}} + \bar{\mathbf{X}} \bar{\mathbf{X}}^T)^{-1} \bar{\mathbf{X}} \sqrt{\frac{\pi}{\nu}} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})}}} \\
&= \frac{(\mathbf{S}_{\mathbf{X}} + \bar{\mathbf{X}} \bar{\mathbf{X}}^T)^{-1} \bar{\mathbf{X}}}{\sqrt{\frac{\nu-2}{\nu} \left[\left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \frac{\nu-2}{\pi} - \bar{\mathbf{X}}^T (\mathbf{S}_{\mathbf{X}} + \bar{\mathbf{X}} \bar{\mathbf{X}}^T)^{-1} \bar{\mathbf{X}} \right]}}.
\end{aligned} \tag{1.5.5}$$

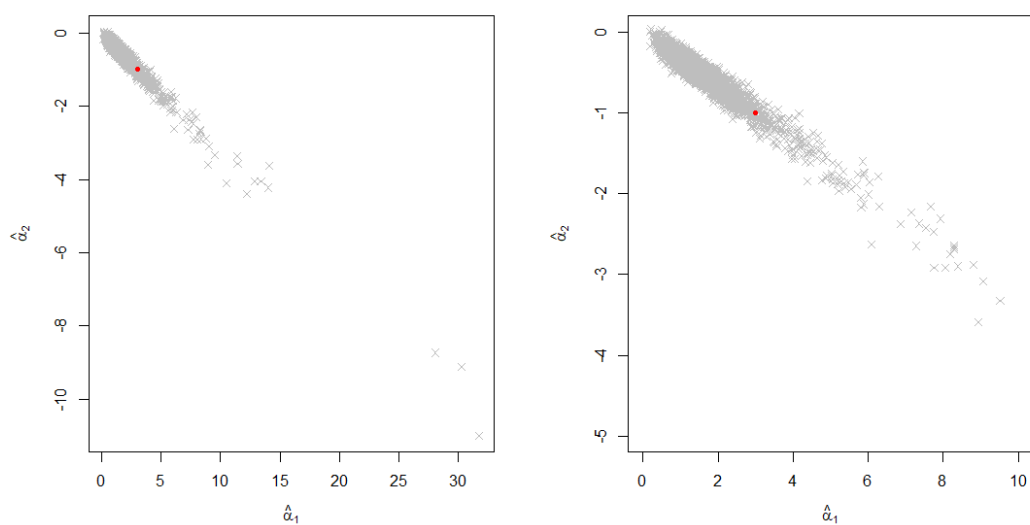
1.5.2 Parameetrite hinnangute käitumine

Järgnevalt uurime saadud teoreetiliste hinnangute (1.5.5) ning (1.5.4) käitumist simuleerimiseksperiment. Simuleerime 10000 erinevat valimit mahuga $n = 1000$ jaotusest $St_{2,4}(\mathbf{\Sigma}, \mathbf{\alpha})$, kus

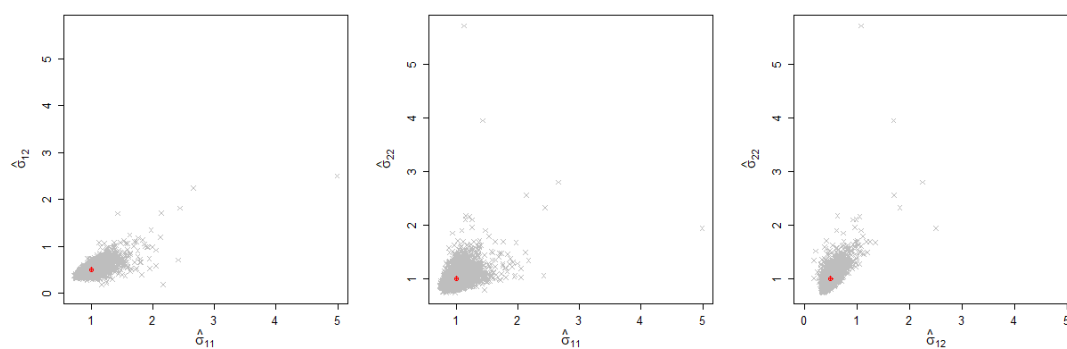
$$\mathbf{\alpha} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}, \quad \nu = 4.$$

Simuleerimisülesande läbiviimisel on kasutatud statistikaprogrammis R funktsiooni *rmst*, mis sisaldub paketis *sn*. Kasutatud programmikood on välja toodud Lisas 1. Iga valimi korral arvutame asümmeetriavektorile $\mathbf{\alpha}$ hinnangud $\hat{\alpha}_1$, $\hat{\alpha}_2$, kasutades võrdust (1.5.5) ning $\mathbf{\Sigma}$ hinnangud $\hat{\sigma}_{11}$, $\hat{\sigma}_{12}$, $\hat{\sigma}_{22}$ kasutades võrdust (1.5.4).

Joonisel 6 on kujutatud parameetrite $\hat{\alpha}_1$ ja $\hat{\alpha}_2$ hinnangud. Punane täpp graafikul tähistab etteantud $\mathbf{\alpha}$ väärtust. Vasakpoolsest graafikult näeme, et enamus hinnatud parameetri väärtuseid jääb esialge $\mathbf{\alpha}$ ümbrusesse. Kuid silma jääb ka 5 väga suure erinevusega hinnangut. Parempoolsest graafikult on lähemalt näha hajuvus algse $\mathbf{\alpha}$ ümber. Joonisel 7 on välja toodud parameetri $\hat{\Sigma}$ väärtuste $\hat{\sigma}_{11}$ ja $\hat{\sigma}_{12}$, $\hat{\sigma}_{11}$ ja $\hat{\sigma}_{22}$ ning $\hat{\sigma}_{12}$ ja $\hat{\sigma}_{22}$ vahelised hajuvusgraafikud. Taaskord tähistab punane täpp graafikul etteantud väärtuse asukohta. Visuaalse uurimise põhjal võime öelda, et $\mathbf{\alpha}$ koordinaatide hinnangud on ebastabiilsemad, kui $\mathbf{\Sigma}$ hinnangute käitumine.



Joonis 6: Simuleeritud valimitelt hinnatud α_1 ja α_2 hajuvuse graafikud.



Joonis 7: Simuleeritud valimitelt hinnatud σ_{11} ja σ_{12} , σ_{11} ja σ_{22} ning σ_{12} ja σ_{22} hajuvuse graafikud.

2 Koopula mõiste

Koopulafunktsioon on mitmemõõtmeline jaotusfunktsioon, mis sisaldab infot marginaalsete tõenäosusjaotustega juhuslike suuruste vahelise sõltuvuse kohta. Sealjuures on koopula invariantne monotoonsete teisenduste suhtes. See annab võimaluse konstrueerida mitmemõõtmelisi jaotusi rakendades andmetele monotoonseid teisendusi [5]. Koopulad on viimase 20 aasta ühed enam uuritud ja kasutatud jaotusfunktsioonid, millest on saanud oluline uus töövahend näiteks finantsmatemaatikas turgude, riskitegurite ja muude uuritavate suuruste modelleerimiseks. Tänu koopulate teooria rakendamisele on võimalik konstrueerida etteantud sõltuvusstruktuuride ja erinevat tüüpi marginaaljaotuste korral mitmemõõtmelisi mudeleid.

Järgnev ülevaade on refereeritud artiklist [4]. Koopulate ajalugu algab prantsuse matemaatiku Maurice Rene Fréchet'ga, kes uuris probleemi, mille siinkohal toome välja kahemõõtmelisel juhul. Olgu meil kaks juhuslikku suurust X_1 ja X_2 jaotusfunktsioonidega vastavalt F_1 ja F_2 , olgu need määratud samal tõenäosusruumil $(\Omega, \mathcal{F}, \mathbb{P})$. Siis mida saame väita kahemõõtmelise jaotusfunktsiooni hulga $\Gamma(F_1, F_2)$ kohta, mille marginaaljaotused on F_1 ja F_2 ? On selge, et hulk $\Gamma(F_1, F_2)$, nn. F_1 ja F_2 Fréchet' klass, ei ole tühi hulk, kuna kui F_1 ning F_2 on sõltumatud, siis $F(x_1, x_2) = F_1(x_1)F_2(x_2)$ kuulub alati hulka $\Gamma(F_1, F_2)$. Küll aga ei ole selge, millised teised elemendid on hulgas $\Gamma(F_1, F_2)$.

Esialgsed uurimused selle probleemi kohta tehti juba 1956. aastal, nii M. R. Fréchet' kui ka E. J. Gumbeli poolt. Küll aga esitas 1959. aastal Abe Sklar mõiste *koopula* mõõduteoorias ning avaldades teoreemi, mis nüüdseks kannab tema nime. Teoreemi tõestust koheselt aga A. Sklar välja ei toonud, esitatud olid vaid mõningad ideed. Aastal 1983 ilmus A. Sklari ja B. Schweizeri koostöös esimene raamat, mis esitas põhilise idee ja tulemused koopulatest. Pärast seda arenes koopulate teooria tõenäosusteoorias kiiresti. 1990. aastal korraldas Dall' Aglio esimese koopulatele pühendatud konverentsi "*Probability distributions with given marginals*". Konverents oli väga edukas ja aitas valdkonna edendamisele oluliselt kaasa.

Suurenenud huvi üheks põhjuseks oli koopula mõiste kasutamine andmeanalüüsis ning rakendamisvõimaluste leidmine rakendusteadustes, näiteks finantsmatemaatikas. Ka hüdroloogias võeti kasutusele koopulad, kuna need võimaldasid luua paindlikumaid mitmemõõtmelisi mudeleid. Nüüdseks on koopulatel kasutusalasid palju ning neid kõiki on pea võimatu üles loendada.

2.1 Peamised mõisted

Esitame mõned peamised definitsioonid mõistmaks koopulate olemust. Olgu p naturaalarv ning märgime ühiklõigu $[0, 1]$ sümboliga \mathbb{I} ning laiendatud reaalsirge $\overline{\mathbb{R}} = [-\infty, +\infty]$.

Definitsioon 2.1.1. Olgu juhuslik vektor $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ tõenäosusruumis $(\Omega, \mathcal{F}, \mathbb{P})$. Siis tema **jaotusfunktsioon** $F_{\mathbf{X}} : \overline{\mathbb{R}}^p \rightarrow \mathbb{I}$ on defineeritud järgmiselt:

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_p) = \mathbb{P} \left(\bigcap_{i=1}^p \{X_i \leq x_i\} \right)$$

Samuti:

(1) $F_{\mathbf{X}}(x_1, x_2, \dots, x_p) = 0$, kui vähemalt ühe argumendi väärtus on $-\infty$,

(2) $F_{\mathbf{X}}(+\infty, +\infty, \dots, +\infty) = 1$.

Definitsioon 2.1.2. Olgu $p \geq 2$ ja F p -dimensionaalne jaotusfunktsioon. Olgu $\sigma = \{j_1, \dots, j_m\}$ hulga $\{1, 2, \dots, p\}$ alamhulk, $1 \leq m < p$. Jaotusfunktsiooni F **σ -marginaaljaotuseks** nimetame jaotusfunktsiooni $F_{\sigma} : \overline{\mathbb{R}}^m \rightarrow \mathbb{I}$, seades jaotusfunktsiooni $p - m$ argumendi väärtuseks $+\infty$. Nimelt iga $x_1, \dots, x_m \in \overline{\mathbb{R}}$ korral

$$F_{\sigma}(x_1, \dots, x_m) = F(y_1, \dots, y_p),$$

kus iga $j \in \{1, 2, \dots, p\}$ korral $y_j = x_j$, kui $j \in \sigma$ ning $y_i = +\infty$ muul juhul.

Märgime ära, et kui juhuslikud suurused X_1, X_2, \dots, X_p on sõltumatud ning kui F_i on X_i jaotusfunktsioon, siis p -dimensionaalse juhusliku vektori $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ jaotusfunktsioon on korrutis:

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_p) = \prod_{i=1}^p F_i(x_i).$$

Definitsioon 2.1.3. Olgu $p \geq 2$. **p -dimensionaalseks koopulaks**, lühidalt p -koopulaks, nimetatakse p -mõõtmelist jaotusfunktsiooni kuubis \mathbb{I}^p , mille ühemõõtmelised marginaaljaotused on ühtlase jaotusega ühiklõigus \mathbb{I} . Märgime

$$C(u_1, u_2, \dots, u_p) : \mathbb{I}^p \rightarrow \mathbb{I}.$$

Seega, iga p -koopula on seotud juhusliku vektoriga $\mathbf{U} = (U_1, U_2, \dots, U_p)^T$, mis iga $i \in \{1, 2, \dots, p\}$ korral on ühtlase jaotusega, $U_i \sim U(0, 1)$. Vastupidiselt kehtib ka,

et iga juhuslik vektor, mille komponendid on ühtlase jaotusega lõigus \mathbb{I} , on teatav koopula. Kuna koopula näol on tegemist mitmemõõtmelise jaotusfunktsiooniga, siis kehtivad ka koopulate korral tähtsamad jaotusfunktsiooni omadused.

2.2 Sklari teoreem

Nagu eespool mainisime on Sklari teoreem koopulate teooria alustalaks. Käesolev paragrahv on refereeritud artiklist [4].

Selleks, et anda parem ülevaade teoreemi väitele, teeme kõigepealt läbi tõestuse kahemõõtmelisel juhul. Olgu meil pidevad juhuslikud suurused X ja Y ning neile vastavad rangelt kasvavad jaotusfunktsioonid F_1 ning F_2 . Rakendades mõlema juhusliku suuruse korral teisendusena nende jaotusfunktsiooni, saame vastavalt $F_1(X) = U_1$ ning $F_2(Y) = U_2$, kus $U_1, U_2 \sim U(0, 1)$. Kuna F_1 ja F_2 on rangelt kasvavad, siis kehtib ka vastupidine võrdus, et $F_1^{-1}(U_1) = X$ ning $F_2^{-1}(U_2) = Y$. Seega kasutades jaotusfunktsiooni definitsiooni $F_X(x) := P(X \leq x)$, saame esitada juhuslike suuruste X ja Y ühisjaotuse jaotusfunktsiooni koopula kaudu:

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) = P(F_1(X) \leq F_1(x), F_2(Y) \leq F_2(y)) \\ &= P(U_1 \leq F_1(x), U_2 \leq F_2(y)) = C(F_1(x), F_2(y)). \end{aligned}$$

Teoreem 2.2.1. *Olgu F p -mõõtmeline jaotusfunktsioon ühemõõtmeliste marginaaljaotustega F_1, F_2, \dots, F_p . Tähistagu A_j funktsiooni F_j väärtuste piirkonda, kus $j = 1, 2, \dots, p$. Siis leidub koopula C nii, et iga $(x_1, x_2, \dots, x_p) \in \mathbb{R}^p$ korral*

$$F(x_1, x_2, \dots, x_p) = C(F_1(x_1), F_2(x_2), \dots, F_p(x_p)). \quad (2.2.1)$$

Selline koopula C on üheselt määratud hulgal $A_1 \times A_2 \times \dots \times A_p$ ning seega ühene kuubis \mathbb{I}^p , kui F_1, F_2, \dots, F_p on kõik pidevad monotoonselt kasvavad jaotusfunktsioonid.

Sklari teoreem esitati esmakordselt juba aastal 1959, kuid teoreemi tõestus kahemõõtmelisel juhul avalikustati alles 1974. aastal A. Sklari ning B. Schweizeri poolt. Teoreemist 2.2.1 järeldub ka järgnev vastupidine implikatsioon, mis on tihti väga oluline just statistiliste mudelite loomisel, võttes eraldi arvesse juhusliku vektori komponentide ühemõõtmelist käitumist ning nende vahelist sõltuvust.

Teoreem 2.2.2. *Kui F_1, F_2, \dots, F_p on ühemõõtmelised pidevad ning rangelt monotoonselt kasvavad jaotusfunktsioonid ja C on suvaline p -koopula, siis funktsioon*

$F : \mathbb{R}^p \rightarrow \mathbb{I}$ kujul (2.2.1) on p -mõõtmeline jaotusfunktsioon marginaaljaotustega F_1, F_2, \dots, F_p .

Seega iga p -mõõtmelist jaotusfunktsiooni on võimalik esitada koopulana C võrduse (2.2.1) kaudu. Kui F_i on pidev iga $i \in \{1, 2, \dots, p\}$ korral, siis saame C kujul

$$C(u_1, u_2, \dots, u_p) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_p^{-1}(u_p)), \quad (2.2.2)$$

kus F_i^{-1} märgib F_i pöördfunktsiooni kujul $F_i^{-1} = \inf\{t | F_i(t) \geq s\}$. Seega, koopulad on sisuliselt viis juhusliku vektori $(X_1, X_2, \dots, X_p)^T$ teisendamiseks juhuslikuks vektoriks $(U_1, U_2, \dots, U_p)^T = (F_1(X_1), F_2(X_2), \dots, F_p(X_p))^T$, kus marginaaljaotused on ühtlase jaotusega ühiklõigus \mathbb{I} , samas säilitatakse monotoonne sõltuvus komponentide vahel.

2.3 Asümmeetriline $t_{p,\nu}$ -koopula

Antud paragrahvis moodustame asümmeetrilise koopula, mis põhineb mitmemõõtmelisel asümmeetrilisel $t_{p,\nu}$ -jaotusel, mille tihedusfunktsioon on kujul (1.5.1). Järgnev esitus on refereeritud artiklist [7]. Olgu meil pidevad juhuslikud suurused X_1, X_2, \dots, X_p tihedusfunktsioonidega vastavalt $f_i(x_i) : \mathbb{R} \rightarrow \mathbb{R}$, kus $i = 1, 2, \dots, p$. Olgu nende ühisjaotuse jaotusfunktsioon $F_{\mathbf{X}}(x_1, x_2, \dots, x_p)$ ning tihedusfunktsioon $f_{\mathbf{X}}(x_1, x_2, \dots, x_p)$. Teoreemi 2.2.1 kasutades saame jaotusfunktsiooni $F_{\mathbf{X}}(x_1, \dots, x_p)$ esitada koopula $C(u_1, u_2, \dots, u_p) : \mathbb{I}^p \rightarrow \mathbb{I}$ kaudu, kus $\mathbf{u} = (u_1, u_2, \dots, u_p)^T \in \mathbb{I}^p$:

$$F_{\mathbf{X}}(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p)), \quad (2.3.1)$$

kus $F_i(x_i) = u_i$. Pideva jaotusega juhuslikul vektoril eksisteerib ka tõenäosustihedus. Tähistame koopula tihedusfunktsiooni $c(u_1, \dots, u_p)$. Koopula tiheduse saame leida, võttes koopula $C(u_1, \dots, u_p)$ funktsioonist tuletise:

$$c(u_1, \dots, u_p) = \frac{\partial^p C(u_1, \dots, u_p)}{\partial u_1 \cdot \dots \cdot \partial u_p}.$$

Kasutades võrdust (2.3.1) saame esitada tihedusfunktsiooni $f_{\mathbf{X}}(x_1, x_2, \dots, x_p)$ koopula tihedusfunktsiooni kaudu:

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_p) = c(F_1(x_1), \dots, F_p(x_p)) \times f_1(x_1) \times \dots \times f_p(x_p).$$

Eelnenud võrdusest saame avaldada koopula tihedusfunktsiooni $c(\mathbf{u}) : \mathbb{I}^p \rightarrow \mathbb{R}$ vektori \mathbf{X} tihedusfunktsiooni ja juhuslike suuruste $X_i, i = 1, \dots, p$, marginaalsete tiheduste kaudu:

$$c(\mathbf{u}) = \frac{f_{\mathbf{X}}(F_1^{-1}(u_1), \dots, F_p^{-1}(u_p))}{f_1(F_1^{-1}(u_1)) \cdot \dots \cdot f_p(F_p^{-1}(u_p))}, \quad (2.3.2)$$

kus $F_1(\cdot), \dots, F_p(\cdot) : \mathbb{R} \rightarrow \mathbb{I}$ on ühemõõtmeliste marginaaljaotuste jaotusfunktsioonid ja $f_1(\cdot), \dots, f_p(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ on vastavate marginaaljaotuste tihedusfunktsioonid.

Järgnevalt kasutame võrdusi (2.3.1) ja (2.3.2), et defineerida mitmemõõtmeline asümmeetriline $t_{p,\nu}$ -koopula.

Definitsioon 2.3.1. Koopulat nimetatakse **asümmeetriliseks $t_{p,\nu}$ -koopulaks** $C_{p,\nu}$, kui

$$C_{p,\nu}(u_1, \dots, u_p; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) = G_{p,\nu}(G_{1,\nu}^{-1}(u_1; \mu_1, \sigma_{11}, \alpha_1), \dots, G_{1,\nu}^{-1}(u_p; \mu_p, \sigma_{pp}, \alpha_p); \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}), \quad (2.3.3)$$

kus $G_{1,\nu}^{-1}(u_i; \mu_i, \sigma_{ii}, \alpha_i) : \mathbb{R}^1 \rightarrow \mathbb{I}, i \in \{1, \dots, p\}$ on ühemõõtmelise $t_{1,\nu}$ -jaotuse jaotusfunktsiooni pöördfunktsioon ja $G_{p,\nu}$ on p -mõõtmelise asümmeetrilise $t_{p,\nu}$ -jaotuse jaotusfunktsioon, mille tihedusfunktsioon on kujul (1.5.1).

Asümmeetrilise $t_{p,\nu}$ -koopula tihedus on

$$c_{p,\nu}(\mathbf{u}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) = \frac{g_{p,\nu}[\{G_{1,\nu}^{-1}(u_1; \mu_1, \sigma_{11}, \alpha_1), \dots, G_{1,\nu}^{-1}(u_p; \mu_p, \sigma_{pp}, \alpha_p)\}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}]}{\prod_{i=1}^p g_{1,\nu}[G_{1,\nu}^{-1}(u_i; \mu_i, \sigma_{ii}, \alpha_i); \mu_i, \sigma_{ii}, \alpha_i]}, \quad (2.3.4)$$

kus $g_{p,\nu}$ on asümmeetrilise $t_{p,\nu}$ -jaotuse tihedusfunktsioon (1.5.1) ning funktsioonid $G_{1,\nu}^{-1}$ on nagu definitsioonis 2.3.1. Tihedusfunktsiooni (2.3.4) lugejas on asümmeetrilise mitmemõõtmelise $t_{p,\nu}$ -jaotuse tihedus. Seega on saadud koopula asümmeetrilise mitmemõõtmelise jaotusega.

2.4 Korrelatsioon

Koopula komponendid võivad olla omavahel sõltuvad juhuslikud suurused. Nende vahelist sõltuvust saame hinnata astakkorrelatsioonikordajate abil, kuna need on invariantset monotoonsete teisenduste suhtes. Levinumad astakkorrelatsioonikordajad on Kendalli τ ning Spearmani ρ_S ning need kordajad defineeritakse koopula

$C(u, v)$ kaudu järgmiselt [5]:

$$\begin{aligned}\tau &= 4 \int \int_{\mathbb{I}^2} C(u, v) dC(u, v) - 1, \\ \rho_S &= 12 \int \int_{\mathbb{I}^2} C(u, v) dudv - 3.\end{aligned}$$

Lineaarse ehk Pearsoni korrelatsioonikordaja ρ korral tekib koopula elementidevahelise korrelatsiooni hindamisel probleeme. Esiteks seetõttu, et ρ ei ole invariantne monotoonsete teisenduste suhtes. See tähendab, et lineaarne korrelatsioon X_i ja X_j vahel ei ole enamasti sama, mis $\log(X_i)$ ja $\log(X_j)$ vahel. 2003. aastal jõudsid F. Lindskog jt artiklis [10] tähtsa tulemuseni ning tõestasid Kendalli τ ning Pearsoni ρ vahelise seose elliptiliste jaotuste korral. Varasemalt oli seos tõestatud vaid kahemõõtmelise normaaljaotuse korral, kuid artiklis näidati, et seos kehtib kõikide pidevate elliptiliste jaotuste korra.

Definitsioon 2.4.1. Olgu $\mathbf{X} = (X_1, \dots, X_p)^T \sim E_p(\mu, \mathbf{V})$ pideva p -mõõtmelise elliptilise jaotusega. Kui tunnuste X_i ja X_j dispersioonid DX_i, DX_j on lõplikud, siis nimetame tunnuste X_i ja X_j vaheliseks **lineaarseks korrelatsioonikordajaks** suurust

$$\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sqrt{DX_i DX_j}}.$$

Definitsioon 2.4.2. Kendalli τ juhuslike suuruste X_i ja X_j korral on defineeritud järgenvalt

$$\tau(X_i, X_j) = P[(X_1 - \tilde{X}_i)(X_2 - \tilde{X}_j) > 0] - P[(X_1 - \tilde{X}_i)(X_2 - \tilde{X}_j) < 0],$$

kus $(\tilde{X}_1, \tilde{X}_2)$ on (X_i, X_j) sõltumatu koopla.

Kasutades neid definitsioone on F. Lindskog jt tõestanud artiklis [10] järgmise tulemuse.

Teoreem 2.4.1. Olgu $\mathbf{X} \sim E_p(\mu, \mathbf{V})$ pideva elliptilise jaotusega. Siis kehtib seos

$$\tau_{ij} = \frac{2}{\pi} \arcsin \rho_{ij}. \quad (2.4.1)$$

3 Kahjude analüüs

3.1 Andmestiku kirjeldus

Antud magistritöö praktilises osas on kasutatud Läti kindlustusseltsi liikluskindlustuse kahjude andmeid. Analüüsimisel kasutatav andmestik sisaldab kolme tunnust - liikluskahjust teavitamise kuupäev, kahjusumma väljamaksmise kuupäev ning kahjusumma suurus. Kahjusumma suurus on esitatud Läti lattides. Liikluskahjust teavitamise kuupäevad jäävad vahemikku 01.01.2001-31.12.2002. Selle perioodi vältel on vaadeldavas kindlustusseltsis esinenud 7519 erinevat kahjujuhtumit. Kolmel real on sisestusviga, kuna kahjusumma väljamaksmise kuupäev on varasem kui kahjust teavitamise kuupäev. Need andmed on analüüsimisel kõrvale jäetud ning kasutame 7516 kahjunõude andmeid. Kõige rohkem on kahjujuhtumitest teavitatud 19. detsembril 2001. Sel päeval oli juhtumeid lausa 61. Samuti on sama aasta 21. ning 28. detsember olnud kõrge teavitamiste arvuga - vastavalt 39 ja 34 kahjujuhtumi teavitust. 2002. aastal on kõige enam kahjujuhtumitest teavitatud 24. mail, kokku 34 teavitust.

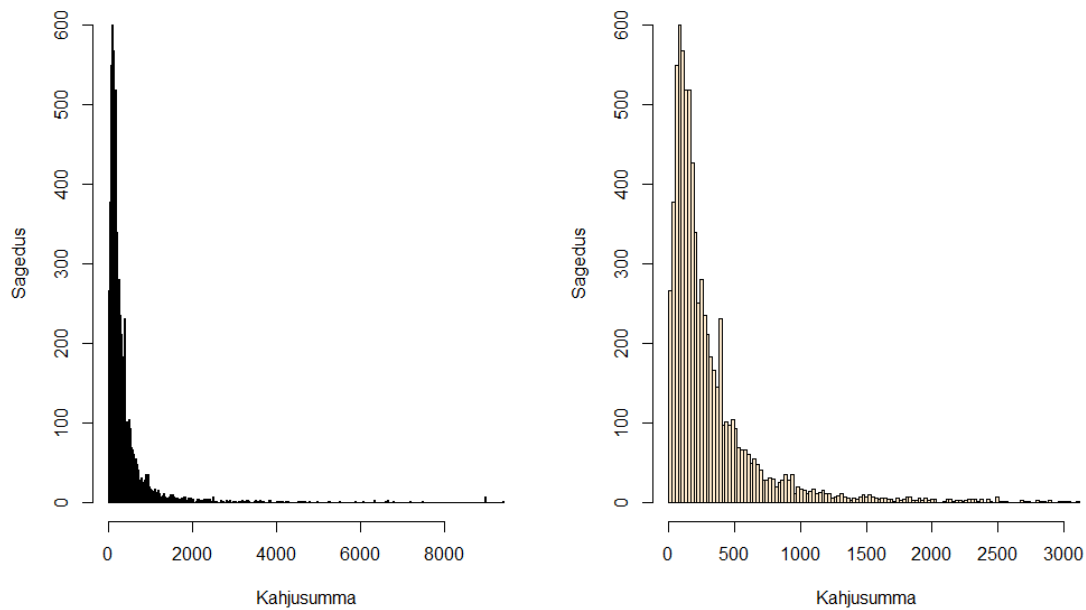
Antud töös kasutame tunnustena kahjusumma suurust ning kahjujuhtumi menetlemise aega, mis kujutab endas päevade arvu kahjujuhtumi teavitamisest kuni viimase kahjusumma osamakse väljamaksmiseni. Kui kindlustusjuhtumi teavitamise ning kahjunõude väljamaksmise kuupäev on sama, on kahjumenetlemise aeg 1 päev. Tabelis 1 on välja toodud mõlema vaadeldava tunnuse empiirilise jaotuse karakteristikud. Mõlemal juhul on asümmeetriakordaja positiivne ning üsna suure väärtusega. See viitab asjaolule, et tegemist on parempoolsete sabadega jaotustega ehk jaotuse mediaan on väiksem keskvaärtusest. Jaotuse standardhälve on väga suur mõlemal juhul, seega tunnuste väärtused on enamjaolt keskmisest väärtusest eemal ning hajuvus on suur.

Karakteristik	Kahjusumma suurus	Kahju menetluse aeg
Keskvaärtus	340,2	81,5
Standardhälve	577	123
Asümmeetriakordaja	7,51	3,51
Vähim väärtus	0,75	1
Suurim väärtus	9432	1142

Tabel 1: Empiiriliste jaotuste karakteristikud.

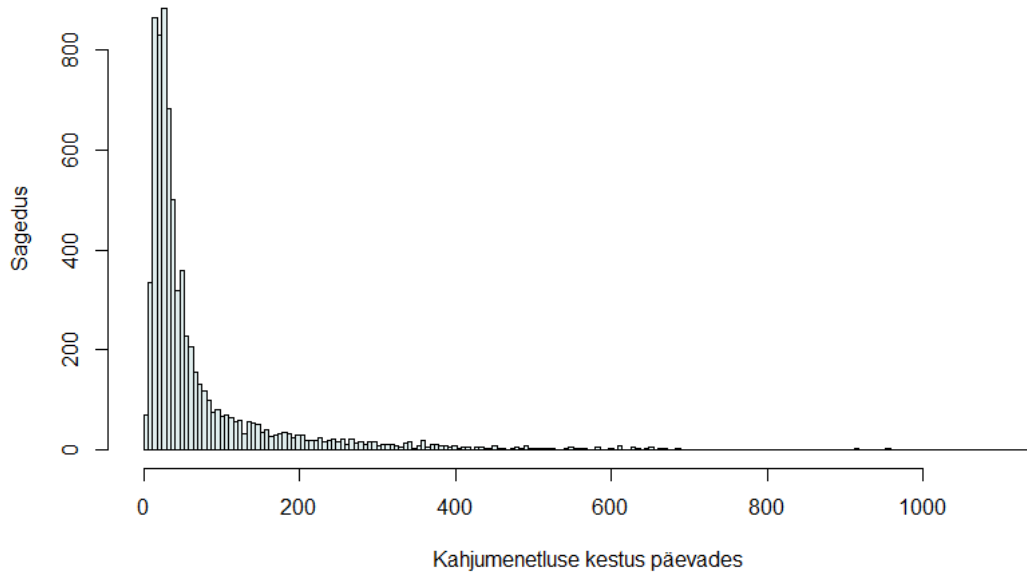
Kõige väiksem väljamakstud kahjusumma on 0,75 latti ning suurim väljamakstud kahjunõue on 9432 latti. Keskmise kahjunõude suurus on 340,29 latti. Joonisel 8

on kujutatud kahjusummade histogrammid. Vasakpoolsel histogrammil on näha kahjusummad kõigi andmete ulatuses, parempoolsel graafikul on välja toodud vaid kahjusummad, mis on väiksemad kui 3000 latti. Näeme, et suurem osa väljamaksest on väikesed ning suuremaid kahjunõudeid esineb harva, mis venitavad jaotuse saba väga pikaks. Selline andmete käitumine on kindlustusandmetele väga tüüpiline. Edaspidi on joonistel empiirilise jaotuse kujutamisel kasutatud kuni 3000 lati suuruseid kahjusummasid, jaotuste sobitamisel on kasutatud kogu andmestikku.



Joonis 8: Kahjusummade histogrammid.

Kahjunõude menetlemise aeg on kaheksal juhul 1 päev ehk et kahjunõue maksti välja kohe samal päeval. 14 juhul maksti kahjunõue välja järgmisel päeval. Kõige pikem kahjumenetlus kestis 1142 päeva. Keskmine kahjumenetluse periood on 81,5 päeva. Kõige enam ehk 189 korral said juhtumid lahenduse 15 päeva jooksul. Joonisel 9 on näha kahjumenetlusele kulunud aja histogramm. Näeme, et enamus kahjujuhtumeid menetletakse päris kiiresti, kuid mõned erijuhud venitavad jaotuse saba samuti pikaks.



Joonis 9: Kahjumenetluse kestuse histogramm.

3.2 Jaotuste sobitamine andmetele

Kaugem eesmärk on leida tunnuste kahemõõtmeline ühisjaotus kasutades asümmeetrilise $t_{p,\nu}$ -jaotuse koopulat. Esmalt proovime sobitada kahele tunnusele praktikas laialdaselt kasutatavat nelja jaotust, et leida kõige paremini andmeid kirjeldav mudel. Mõlemal juhul kasutame modelleerimisel gamma-, Pareto, Weibulli ning lognormaalset jaotust.

Definitsioon 3.2.1. *Mittenegatiivne juhuslik suurus X on **gammajaotusega** kujuparameetriga α ja skaalaparameetriga β , tähistame $X \sim \Gamma(\alpha, \beta)$, kui tema tihedusfunktsioon on kujul*

$$f_X(x; \alpha, \beta) = \begin{cases} x^{\alpha-1} e^{-\frac{x}{\beta}} \frac{1}{\beta^\alpha \Gamma(\alpha)}, & \text{kui } x > 0, \\ 0, & \text{mujal,} \end{cases}$$

kus $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ on gammafunktsioon.

Definitsioon 3.2.2. *Juhuslik suurus X on **Weibulli jaotusega** kujuparameetriga k ja skaalaparameetriga λ , tähistame $X \sim We(k, \lambda)$, kui tema tihedusfunktsioon*

avaldub kujul

$$f_X(x; k, \lambda) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, & \text{kui } x > 0, \\ 0, & \text{mujal.} \end{cases}$$

Definitsioon 3.2.3. Juhuslik suurus X on **Pareto jaotusega** kujuparameetriga k ja skaalaparameetriga λ , tähistame $X \sim Pa(k, \lambda)$, kui tema tihedusfunktsioon avaldub kujul

$$f_X(x; k, \lambda) = \begin{cases} \frac{k\lambda^k}{(x+\lambda)^{k+1}}, & \text{kui } x > 0, \\ 0, & \text{mujal.} \end{cases}$$

Definitsioon 3.2.4. Juhuslik suurus X on **lognormaalse jaotusega** asukohaparameetriga α ja skaalaparameetriga σ , tähistame $X \sim LN(\alpha, \sigma)$, kui tema tihedusfunktsioon avaldub kujul

$$f_X(x; k, \lambda) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, & \text{kui } x > 0, \\ 0, & \text{mujal.} \end{cases}$$

Gamma-, Weibulli, Pareto ning lognormaalse jaotuse parameetrite suurima tõepära meetodil hindamiseks on statistikaprogrammis R kasutatud funktsiooni *fitdist*, mis sisaldub paketis *fitdistrplus*.

3.2.1 Kahjusumma suurus

Joonisel 10 on kujutatud kahjusumma suuruse empiirilisele jaotusele sobitatud jaotuste tihedusfunktsioonid. Visuaalsel vaatlusel näib, et kõige täpsemalt kirjeldab tunnuse empiirilist jaotust lognormaalne jaotus. Teised sobitatud jaotused ei kirjelda nii täpselt väiksemaid kahjusid. Tabelis 2 on välja toodud sobitatud jaotuste parameetrite suurima tõepära hinnangud.

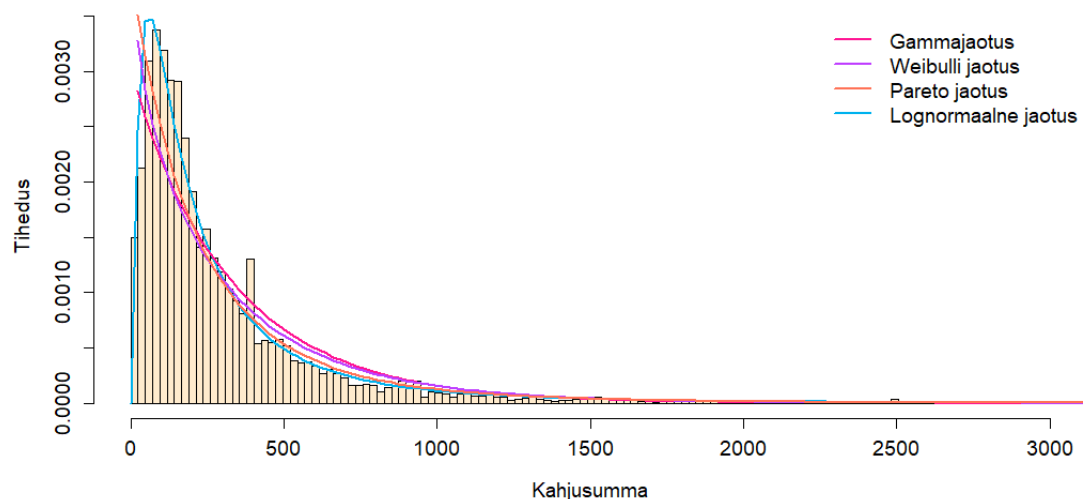
Selleks, et hinnata jaotuste sobivust empiirilise jaotusega, kasutame Kolmogorov-Smirnovi kooskõlatesti ning lisaks leiame ka Akaike ja Schwarzi informatsiooni-kriteeriumite väärtused, tähistame vastavalt *AIC* ja *BIC*. Kolmogorov-Smirnovi kooskõlatesti korral märgime teststatistiku väärtuse vastavalt D , kriitiline väärtus valimi mahu 7516 korral on 0,01603. Informatsiooni-kriteeriumite *AIC* ja *BIC* väärtused kirjeldavad sobivust logaritmilise tõepära kaudu võttes arvesse ka hinnatavate parameetrite arvu jaotuses. Mõlema kriteeriumi korral on paremini sobiv jaotus see, mille korral kriteeriumi väärtus on väiksem.

$$AIC = 2k - 2\ln L,$$

kus k on mudelis hinnatavate parameetrite arv ning L on maksimeeritud logaritmiline tõepära;

$$BIC = \ln(n)k - 2\ln(\hat{L}),$$

kus n on vaatluste arv.



Joonis 10: Kahjusummade empiiriline jaotus ning sobitatud jaotuste tihedusfunktsioonid.

Jaotus	Parameetrite hinnangud
Gamma	$\alpha = 0,976$ $\beta = 2,87 \cdot 10^{-3}$
Weibull	$k = 0,907$ $\lambda = 320,99$
Pareto	$k = 4,31$ $\theta = 1095,48$
Lognormaalne	$\alpha = 5,24$ $\sigma = 1,08$

Tabel 2: Sobitatud jaotuste parameetrite hinnangud.

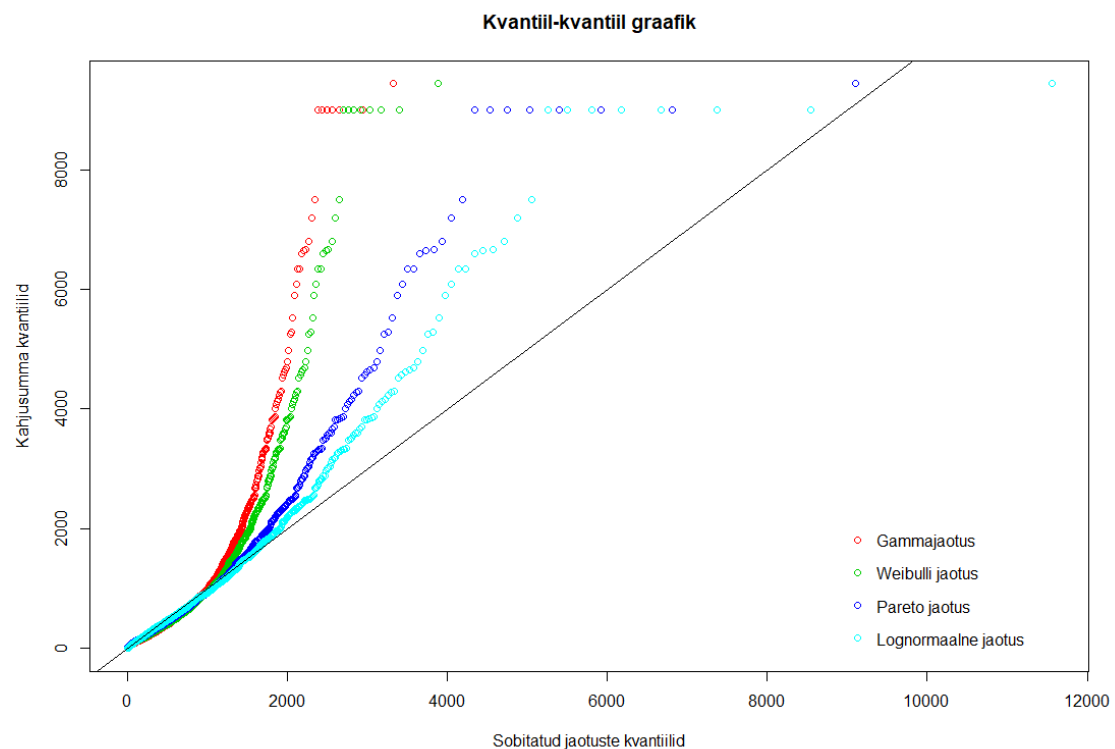
Tabelis 3 on välja toodud Kolmogorov-Smirnovi kooskõlateesti teststatistiku D ning AIC ning BIC väärtused kõigi nelja sobitatud jaotuse korral. Teststatistiku D väärtus on kõigil juhtudel suurem kriitilisest väärtusest. Kõige väiksem on teststatistiku väärtus lognormaalse jaotuse korral - 0,0279. Järgneb Weibulli jaotus, mille korral $D = 0,0774$. Pareto jaotuse korral on teststatistiku väärtus 0,0865 ning suurim on D väärtus Gamma jaotuse korral - 0,0891. Kuna aga meil on tegu üpris suure valimiga ja seega võivad ka väiksemad erinevused empiirilise ja teoreetilise jaotusfunktsiooni väärtustes nullhüpoteesi, et andmed on etteantud teoreetilise jaotusega, ümber lükata. Samuti tuleks arvesse võtta ka asjaolu, et hetkel on Kolmogorov-Smirnovi testid läbi viidud hinnatud parameetrite korral, seesugune lähenemine ei ole aga teoreetiliselt korrektne ning olulisustõenäosuste väärtused on

suuremad tegelikest. Seega saame leitud teststatistikuid kasutada omavaheliseks võrdluseks, küll aga ei saa me teha statistilisi otsuseid jaotuste sobivuse kohta.

Vaadates informatsioonikriteeriume näeme, et lognormaalse jaotuse korral on mõlema informatsioonikriteeriumi väärtus väiksem ning seega võime väita, et lognormaalne jaotus parameetritega $\alpha = 5,25$ ning $\sigma = 1,08$ kirjeldab kõige paremini kahjusummade jaotumist käesolevas andmestikus.

Jaotus	D	AIC	BIC
Gammajaotus	0,0891	102667	102681
Weibulli jaotus	0,0774	102508	102522
Pareto jaotus	0,0865	101778	101792
Lognormaalne jaotus	0,0297	101180	101194

Tabel 3: Kooskõlatesti ning informatsioonikriteeriumite väärtused tunnuse "Kahjusumma" korral.



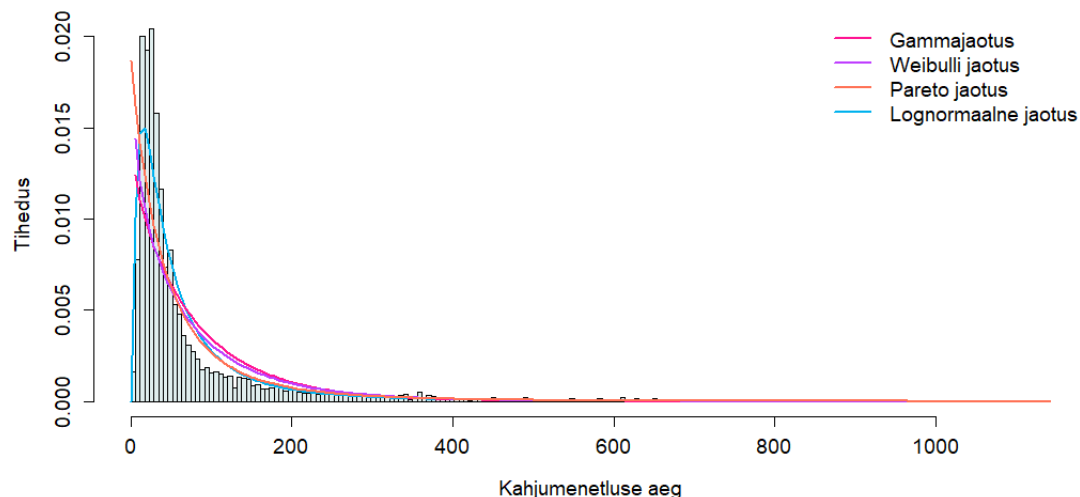
Joonis 11: Kahjusumma andmetele sobitatud jaotuste kvantiil-kvantiil graafikud.

Joonisel 11 on välja toodud sobitatud jaotuste kvantiil-kvantiil graafikud. Graafiku x -teljel on toodud sobitatud jaotuste kvantiilid ning y -teljel valimi kvantiilid. Sobitatud jaotuse hea sobivuse korral peaksid saadud punktid asetsema sirgel $y(x) = x$. On näha, et andmete põhiosas on kõik sobitatud jaotused väga hästi andmete jaotumist kirjeldavad. Küll aga sabaosas jäävad sobitatud jaotused reaalistest andmetest

üpris kaugemale. Suurima tõepära meetodil leitud hinnangud keskenduvadki põhiosa täpsemale hindamisele, kuna suurem osa vaatlusi on seal. Küll aga kinnitab ka kvantiil-kvantiil graafik eelpool tehtud järeldust, et lognormaalne jaotus kirjeldab kahjusumma suuruste jaotumist kõige paremini. Samas viitab graafik sellele, et jaotuse saba oleks mõistlik eraldi modelleerida. Kantiil-kvantiil graafikult näeme, et lognormaalne jaotus ei ole kahjude kirjeldamiseks piisavalt konservatiivne.

3.2.2 Kahjumenetluse aeg

Joonisel 12 on kujutatud kahjumenetluse aja empiirilisele jaotusele sobitatud jaotuste tihedusfunktsioonid. Visuaalsel vaatlusel näib jälle, et kõige täpsemalt kirjeldab tunnuse empiirilist jaotust lognormaalne jaotus ning teised sobitatud jaotused ei kirjelda nii täpselt lühema vältega kahjukäsitluste jaotumist. Tabelis 4 on esitatud sobitatud jaotuste parameetrite suurima tõepära hinnangud. Tabelis 5 välja toodud Kolmogorov-Smirnovi kooskõlateesti teststatistiku D väärtuseid võrreldes näeme, et lognormaalse jaotuse korral on väärtus väikseim. Samuti on lognormaalse jaotuse korral ka informatsioonikriteeriumite AIC ja BIC väärtused väikseimad. Seega sobib lognormaalne jaotus parameetritega $\alpha = 3,77$ ning $\sigma = 1,04$ kõige paremini iseloomustama kahjumenetluse aja empiirilist jaotust.



Joonis 12: Kahjumenetluse aja empiiriline jaotus ning sobitatud jaotuste tihedusfunktsioonid.

Joonisel 13 on kujutatud sobitatud jaotuste kvantiil-kvantiil graafik. Taaskord näeme, et andmete põhiosas kirjeldavad sobitatud jaotused väga hästi andmete jao-

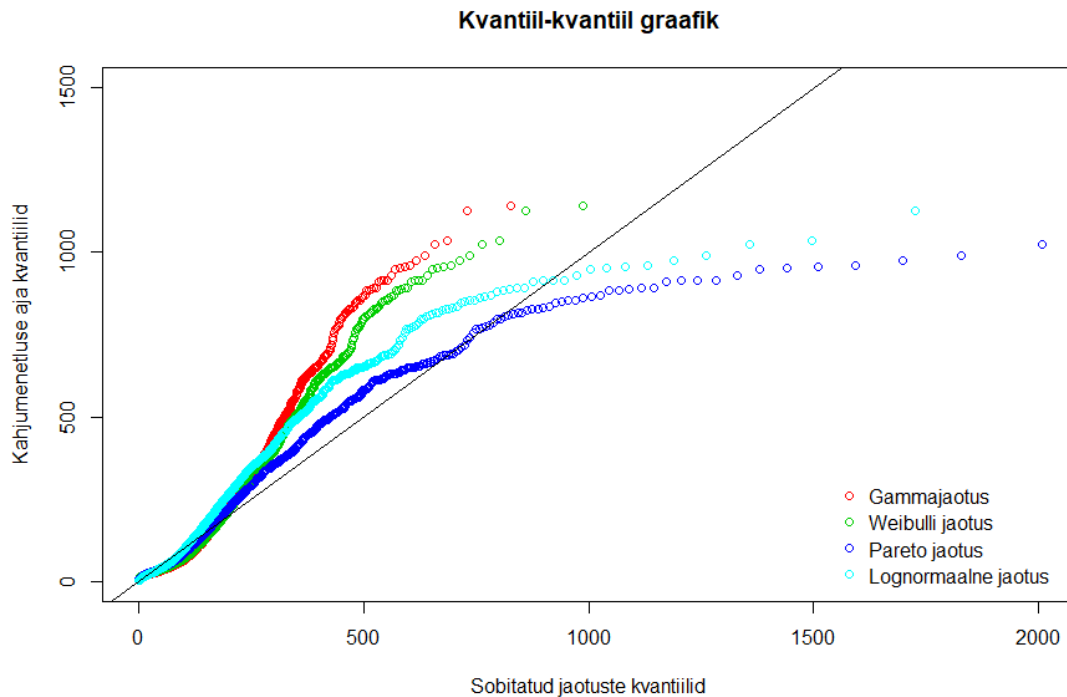
Jaotus	Parameetrite hinnangud	
Gamma	$\alpha = 0,927$	$\beta = 11,4 \cdot 10^{-3}$
Weibull	$k = 0,879$	$\lambda = 75,093$
Pareto	$k = 2,922$	$\theta = 156,67$
Lognormaalne	$\alpha = 3,77$	$\sigma = 1,04$

Tabel 4: Sobitatud jaotuste parameetrite hinnangud.

Jaotus	D	AIC	BIC
Gammajaotus	0,1671	81157	81171
Weibulli jaotus	0,1347	80907	80921
Pareto jaotus	0,1400	80026	80040
Lognormaalne jaotus	0,0861	78621	78635

Tabel 5: Kooskõlatesti ning informatsioonikriteeriumite väärtused tunnuse "Kahjumenetluse aeg" korral.

tumist, küll aga sabaosas jäävad sobitatud jaotused reaalistest andmetest üpris kaugele. Kvantiil-kvantiil graafik kinnitab seda, et lognormaalne jaotus kirjeldab kahjukäsitusaja jaotumist kõige paremini. Samas on erinevalt kahjusummasid kirjeldavast jaotusest tegemist olukorraga, kus suurte väärtuste esinemist ülehinnatakse ja seega on tegemist liiga konservatiivse mudeliga.



Joonis 13: Kahjusumma käsitusaja andmete sobitatud jaotuste kvantiil-kvantiil graafikud.

3.3 Ühisjaotuse leidmine

Järgnevalt proovime moodustada ühisjaotuse tunnustele kahjusumma suurus ning kahjumenetluse aeg. Vastupidiselt asümmeetrilise $t_{p,\nu}$ -koopula tihedus- ning jaotusfunktsioonide kirjapiltide keerukusele, on antud koopula väärtusi üpriski kerge simuleerida.

3.3.1 Algoritm asümmeetrilise $t_{p,\nu}$ -koopula väärtuste simuleerimiseks

Siinkohal paneme kirja sammud ühisjaotuse leidmiseks $t_{p,\nu}$ -koopula abil. Järgnev simulatsioonieeskiri baseerub $t_{p,\nu}$ -koopula simuleerimiseeskirjale ([3], § 6.2) ning asümmeetrilise $t_{p,\nu}$ -jaotuse definitsioonile.

- (1) Leiame maatriksi \mathbf{A} Cholesky dekompositsiooni abil korrelatsioonimaatriksist \mathbf{R} : $\mathbf{A}\mathbf{A}^T = \mathbf{R}$.
- (2) Simuleerime p erinevat väärtust \mathbf{z}_i standardsest normaaljaotusest $N(0, 1)$ ja moodustame p -vektori $\mathbf{z} = (z_1, \dots, z_p)^T$.
- (3) Leiame vektori $\mathbf{x} = \mathbf{A}\mathbf{z}$.
- (4) Simuleerime suuruse Z_0 standardsest normaaljaotusest $N(0, 1)$.
- (5) Leiame asümmeetrilise normaaljaotusega vektori \mathbf{y} realisatsiooni järgmiselt:

$$\mathbf{y} = \begin{cases} \mathbf{x}, & \text{kui } \boldsymbol{\alpha}^T \mathbf{x} > Z_0, \\ -\mathbf{x}, & \text{kui } \boldsymbol{\alpha}^T \mathbf{x} \geq Z_0. \end{cases} \quad (3.3.1)$$

- (6) Simuleerime χ^2 -jaotusest juhusliku suuruse väärtuse V .
- (7) Leiame vektori $\mathbf{t} = \sqrt{\frac{\nu}{V}}\mathbf{y}$.
- (8) Moodustame vektori \mathbf{u} nii, et iga koordinaadi korral $u_i = G_{1,\nu}(t_i; 0, 1, \alpha_i)$, $i \in \{1, \dots, p\}$, kus $G_{1,\nu}$ on ühemõõtmelise asümmeetrilise t_ν -jaotuse jaotusfunktsioon.
- (9) Moodustame vektori $\mathbf{x} = (F_1^{-1}(u_1), \dots, F_p^{-1}(u_p))$, kus $F_i^{-1}(\cdot)$ on esialgse juhusliku suuruse X_i jaotusfunktsiooni pöördfunktsioon.
- (10) Kordame samme 2-9 k korda.

3.3.2 Ühisjaotusest simuleerimine

Esmalt leiame kindlustusandmete punkthinnangud keskväärtsvektorile ning dispersioonimaatriksile:

$$\bar{\mathbf{X}} = \begin{pmatrix} 340,3 \\ 81,5 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 333097,8 & 7141,0 \\ 7141,0 & 15299,2 \end{pmatrix}.$$

Korrelatsioonikaatriksi \mathbf{R} hinnangu saame valemiga $\hat{\mathbf{R}} = \mathbf{S}_d^{-\frac{1}{2}} \mathbf{S} \mathbf{S}_d^{-\frac{1}{2}}$, kus \mathbf{S}_d määrab diagonaliseeritud maatriksit \mathbf{S} :

$$\hat{\mathbf{R}} = \begin{pmatrix} 1 & 0,1 \\ 0,1 & 1 \end{pmatrix}.$$

Kasutades teoreemi 2.4.1 väidet, saame leida Kendalli τ hinnangu järgmiselt

$$\hat{\tau}_{12} = \frac{2}{\pi} \arcsin(\hat{r}_{12}) = \frac{2}{\pi} \arcsin(0,1) = 0,064.$$

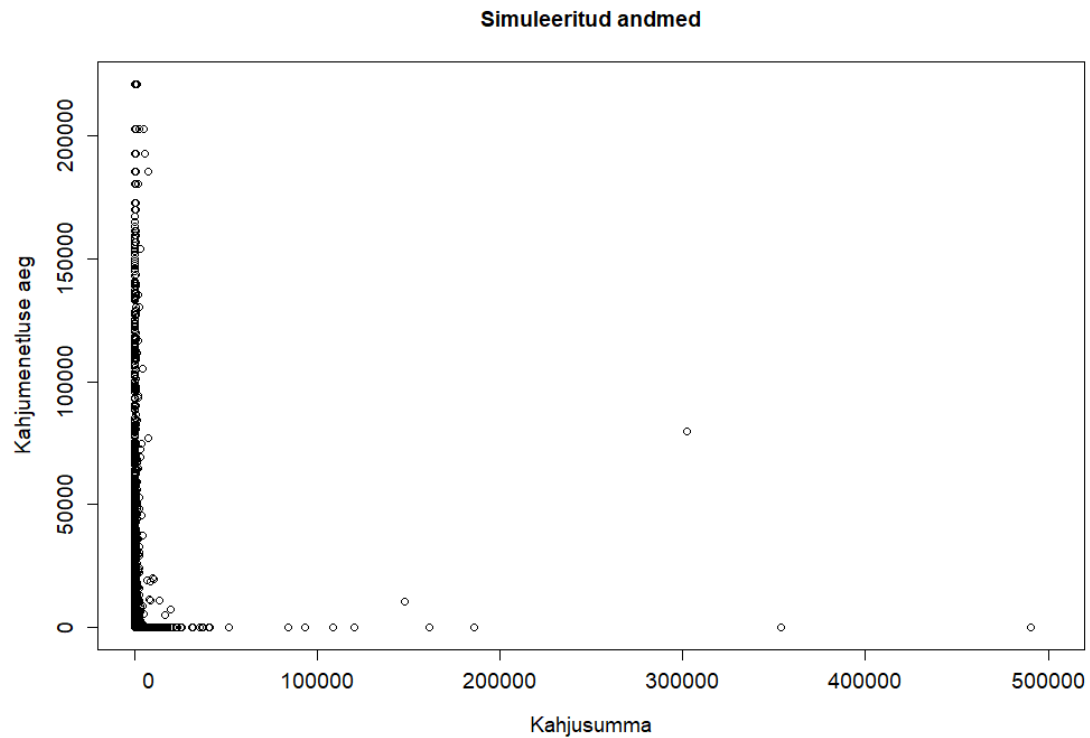
Seega on korrelatsioon nende kahe tunnuse vahel väike. Nüüd leiame hinnangud parameetritele $\boldsymbol{\alpha}$ ja $\boldsymbol{\Sigma}$ kasutades võrdusi (1.5.5), (1.5.4):

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 0,003 \\ 0,014 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 224447,7 & 17437,5 \\ 17437,5 & 10970,8 \end{pmatrix}.$$

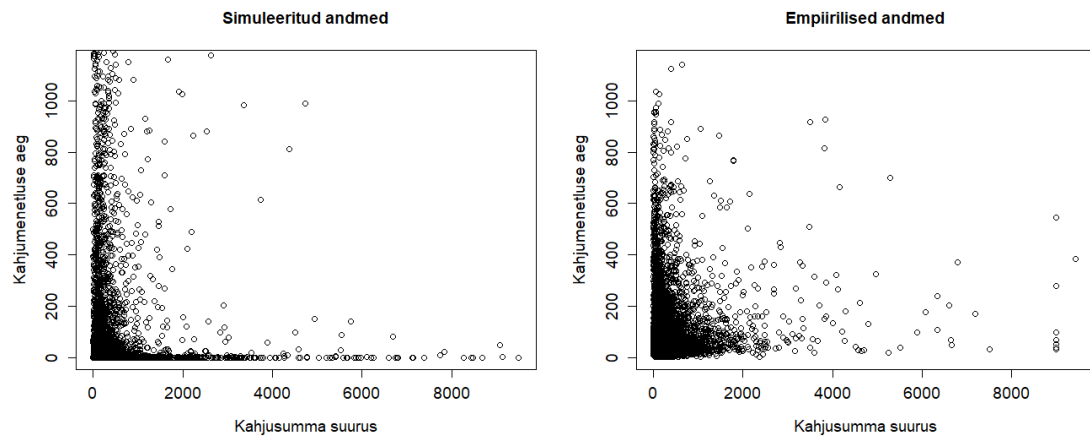
Järgnevalt simuleerime asümmeetrilisest $t_{2,4}$ -koopulast valimi mahuga $k = 7500$ kasutades eelnevalt toodud algoritmi. Simuleerimisel kasutatud algoritmi R kood on välja toodud Lisas 2 koos kahjude analüüsiga. Me moodustame 2-mõõtmelise juhusliku suuruse, kus kahjusumma ehk esimese juhusliku suuruse X_1 marginaaljaotus on LN -jaotus, $X_1 \sim LN(\boldsymbol{\alpha}_1 = 5,25, \boldsymbol{\sigma}_1 = 1,08)$ ning kahjumenetluse aja X_2 marginaaljaotus on $X_2 \sim LN(\boldsymbol{\alpha}_1 = 3,77, \boldsymbol{\sigma}_1 = 1,04)$. Vabadusastmete arv $\nu = 4$. Simuleerimisel kasutame sisendina korrelatsioonimaatriksi hinnangut $\hat{\mathbf{R}}$ ning asümmeetria parameetri hinnangut $\hat{\boldsymbol{\alpha}}$.

Joonisel 14 on näha ühe simuleeritud valimi andmete jaotumine. Näeme, et tunnuse "kahjumenetluse aeg" korral on sabaosas väga palju vaatlusi, mida empiiriliste andmete korral nii suurel määral ei esinenud. Kindlasti üheks mõjutajaks on liialt konservatiivne marginaaljaotus, mida simuleerimisel kasutasime. Joonisel 15 on välja toodud simuleeritud andmete ning empiiriliste andmete graafikud empiiriliste andmete esinemisvahemikus. Simuleeritud andmete korral on näha, et esineb palju vaatlusi, mil nii kahjusumma kui ka kahjumenetluse aeg on väikesed. Kuid

siiski on sabaosades rohkem vaatlusi kui empiirilistel andmetel.



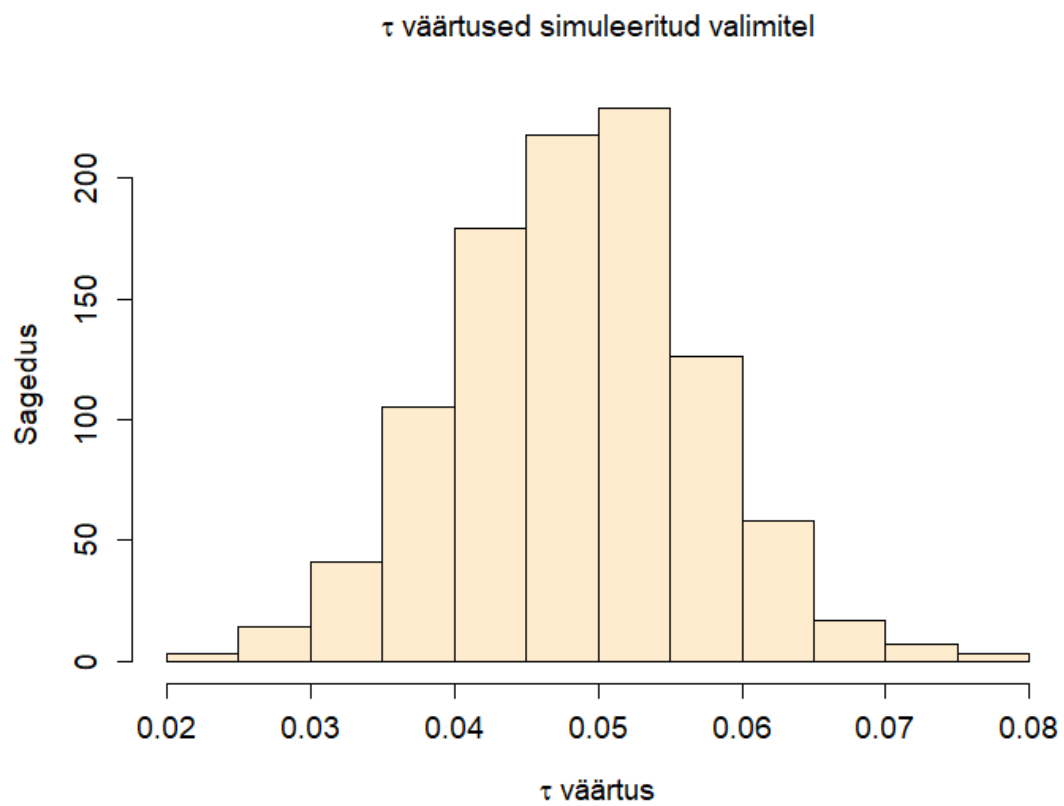
Joonis 14: Simuleeritud ühe valimi väärtused.



Joonis 15: Simuleeritud väärtuste ning esialgsete andmete võrdlus.

Selleks, et hinnata simuleeritud valimites tunnustevahelise sõltuvust, genereerime seesuguseid valimeid 1000 tükki ning leiame iga valimi pealt Kendalli korrelatsioonikordaja τ . Joonisel 16 on τ väärtuste histogramm. Keskmine väärtus on 0,048. Histogrammilt on näha, et ühegi valimi korral ei ole tunnustevaheline sõltuvus suuresti muutunud, alati on korrelatsioon tunnuste vahel olnud väike. Vähim väärtus

on olnud 0,025 ning suurim 0,078.



Joonis 16: Simuleeritud valimite τ väärtused.

Kokkuvõte

Käesolevas magistritöös on tutvustatud asümmeetriliste elliptiliste jaotuste klassi kuuluvat mitmemõõtmelist asümmeetrilist $St_{p,\nu}$ -jaotust ning selle jaotuse baasil konstrueeritud koopulat.

Esimeses peatükis on antud lühiülevaade elliptilistest jaotustest ning mitmemõõtmeliste jaotuste arvkarakteristikutest. Seejärel on tutvustatud asümmeetrilist normaaljaotust ning esitatud jaotuse keskväärtsvektor, dispersioonimaatriks ning hinnangud parameetritele. Lisaks on tutvustatud χ_ν - pöördjaotust ning leitud selle jaotuse momendid. Peale seda on antud $St_{p,\nu}$ -jaotuse definitsioon ning leitud on hinnangud jaotuse parameetritele. Samuti on läbi viidud simuleerimiseksperiment parameetrite hinnangute hajuvuse hindamiseks. Teises peatükis on antud ülevaade koopulate teooriast ning sisse toodud asümmeetriline $t_{p,\nu}$ -koopula.

Kolmandas peatükis on leitud kindlustusandmete empiirilist jaotust kirjeldavad jaotused. Mõlema uuritud tunnuse korral andis parima tulemuse lognormaalne jaotus. Tunnuse "kahjusumma" korral oli sobivaimaks $LN(\alpha = 5, 25, \sigma = 1, 08)$ jaotus, kuid sabaosas olevate vaatluste hindamiseks ei olnud antud jaotus piisavalt konservatiivne ning alahindas suuremate väärtuste esinemist. Tunnuse "kahjumenetluse aeg" korral oli sobivamaiks jaotuseks $LN(\alpha = 3, 77, \sigma = 1, 04)$ jaotus, kuid antud jaotus on selle tunnuse jaoks liialt konservatiivne ning seega suuremate väärtuste esinemist ülehinnatakse. Empiirilise jaotuse paremaks kirjeldamiseks oleks tulnud nähtavalt sobitada eraldi kaks jaotust, põhiosale ning sabaosale eraldi, kuid see ei olnud antud töö eesmärgiks. Andmete genereerimiseks $St_{2,4}$ -koopulast kasutati ühemõõtmeliste marginaaljaotustena leitud LN -jaotusi mõlema tunnuse korral. Võrreldes saadud juhuslikku vektorit esialgsete andmetega selgus, et kooskõla empiiriliste andmetega on hea kuid sabaosas ei suuda antud mudel siiski tunnuste käitumist kõige täpsemini tabada.

Magistritöö eesmärgiks oli anda ülevaade $St_{p,\nu}$ -jaotusest ja selle baasil moodustatud koopulast ning näidata selle rakendatavust ebasümmeetriliste jaotuste modelleerimisel. Autor hindab eesmärgi täidetuks.

Summary

This Master's thesis introduces the multivariate skew $St_{p,\nu}$ -distribution, which belongs to the class of skew elliptical distributions, and a copula which is constructed on the basis of this distribution.

The first chapter gives a brief overview of the elliptical distributions and the characteristics of the multivariate distributions. The skew normal distribution is then introduced, and expressions of the mean vector, the dispersion matrix and their estimates are presented. In addition, the inverse χ_ν -distribution has been introduced, and the moments of this distribution have been derived. Thereafter the definition of $St_{p,\nu}$ -distribution is given, and estimates for the parameters are found. Also, a simulation experiment for estimating dispersion of the parameter estimates is carried out in a two-dimensional case. The second chapter gives an overview of the copula theory and the skew $t_{p,\nu}$ -copula is presented.

In the third chapter, distributions describing the empirical distribution of insurance data were found. Two variables - "claim size" and "claim settlement time" were under consideration. The marginal distribution were approximated by Gamma, Weibull, Pareto and log-normal distribution. For both studied variables, the best fit was obtained by a log-normal distribution. For variable "claim size" $LN(\alpha = 5.25, \sigma = 1.08)$ was the most appropriate distribution, but this distribution was not conservative enough to estimate observations in the tail. For variable "claim settlement time" $LN(\alpha = 3.77, \sigma = 1.04)$ distribution was the most appropriate, but this distribution was too conservative and hence higher possible values were overestimated. For better description of the empirical distribution, it would be possible to find two different distributions to describe the main part and tail area of the distribution, but this was not the purpose of this thesis. However, for the $St_{2,4}$ data simulation, the LN distributions for both variables were used. Compared to the initial data, it was pointed out that the fit between the model and empirical data was good but in the tail area, however, this model was not able to pinpoint the behaviour of the variables most adequately.

The aim of the Master's thesis was to give an overview of the $St_{p,\nu}$ - distribution and the copula formed on the basis of the distribution, also to demonstrate its applicability in modelling skewed distributions. Author finds that the goal was achieved.

Viited

- [1] Azzalini, A., Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **65**, No. 2 (2003), 367-389;
- [2] Azzalini, A., Capitanio, A. (2014) *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge;
- [3] Cherubini, U., Luciano, E., Vecchiato, W. (2004) *Copula Methods in Finance*, Wiley, Chichester;
- [4] Durante, F., Sempi, C. (2010) Chapter 1: Copula Theory: An Introduction. In: Jaworski, P. et.al. (Eds.) *Copula Theory and Its Applications. Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, Springer, Heidelberg, 3-31;
- [5] Joutsu, N. (2014) *Olulise valimi meetod krediidiriski hindamisel koopulatega*, Magistritöö, matemaatilise statistika instituut, Tartu Ülikool;
- [6] Kollo, T., Käärrik, M., Selart, A. (2018) The asymptotic normality of the estimators for a multivariate skew normal distribution, *Communications in Statistics - Theory and Methods*, **47**, 3640-3655;
- [7] Kollo, T., Pettere, G. (2010) Chapter 15: Parameter estimation and application of the multivariate sSkew t -copula, In: Jaworski, P. et.al. *Copula Theory and Its Applications. Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, Springer, Heidelberg, 289-298;
- [8] Kollo, T., von Rosen, D. (2005) *Advanced Multivariate Statistics with Matrices* Springer, Voorburg;
- [9] Kotz, S., Nadarajah, S. (2004) *Multivariate t Distributions and Their Applications*, Cambridge University Press, Cambridge;
- [10] Lindskog F., McNeil A., Schmock U. (2003) Kendall's tau for elliptical distributions. In: Bol G., Nakhaeizadeh G., et.al. (Eds) *Credit Risk. Contributions to Economics.*, Physica-Verlag HD;
- [11] Mitmemõõtmeline statistika. Veebikonspekt, URL (vaadatud: 10.05.2019) https://courses.ms.ut.ee/MTMS.01.065/2018_spring/uploads/Main/mms-loengukonspekt.pdf

- [12] Tiit, E., Parring, A., Möls, T. (1977) *Tõenäosusteooria ja matemaatiline statistika*, Valgus, Tallinn;

Lisad

Lisa 1. Simuleerimisülesande R-i kood

```
# Simuleerime valimi  $St_{-}(2, 4)$  jaotusest, valimi mahuga  $n=1000$ .
# Hindame valimis  $\alpha_1$  ja  $\alpha_2$  ja Sigma maatriksi
# Kordame 10000 korda ja koostame hajuvusgraafikud
library(sn) #skew-jaotused

Sigma=diag(2)
Sigma[2,1]=Sigma[1,2]=0.5
alpha=c(3,-1)
nu=4
hinnangud_Sigma=setNames(data.frame(matrix(ncol = 4, nrow = 0)),
                           c("sigma11", "sigma12", "sigma21", "sigma22"))
hinnangud_alpha=setNames(data.frame(matrix(ncol = 2, nrow = 0)),
                           c("alpha1", "alpha2"))

i=1
while (i <=10000){
  X=rmst(1000, Omega=Sigma, alpha=alpha, nu=nu)
  X_kaetud=matrix(colMeans(X), nrow=2, ncol=1)
  S=cov(X)
  juure_alune=(nu-2/nu)*(((gamma((nu-1)/2)/gamma(nu/2))^2*(nu-2)/pi)-
                        t(X_kaetud)%*%
                        solve(S+X_kaetud%*%t(X_kaetud))%*%X_kaetud)

  if (juure_alune>0){
    Sigma_hinnang=((nu-2)/nu)*(S+X_kaetud%*%t(X_kaetud))
    Sigma_hinnang=c(Sigma_hinnang[,1], Sigma_hinnang[,2])
    nimetaja=as.numeric(1/sqrt(juure_alune))
    alpha_hinnang=nimetaja*(solve(S+X_kaetud%*%t(X_kaetud))%*%X_kaetud)
    hinnangud_Sigma=rbind(hinnangud_Sigma, Sigma_hinnang)
    hinnangud_alpha=rbind(hinnangud_alpha, t(alpha_hinnang))
    i=i+1
  }
}
#hinnangud_Sigma
#hinnangud_alpha

par(mfrow=c(1,2))
plot(hinnangud_alpha, pch=4, col="Gray", xlab=bquote(hat(alpha)[1]),
      ylab=bquote(hat(alpha)[2]))
```

```

points(x=alpha[1], y=alpha[2], pch=20, col="red")
plot(hinnangud_alpha, pch=4, col="Gray", xlab=bquote(hat(alpha)[1]),
      ylab=bquote(hat(alpha)[2]), xlim=c(0, 10), ylim=c(-5, 0))
points(x=alpha[1], y=alpha[2], pch=20, col="red")

par(mfrow=c(1,3), mar= c(5, 5, 2, 2) + 0.1, oma = c(1,1,1,1) + 0.1)
plot(x=hinnangud_Sigma[,1], y=hinnangud_Sigma[,2], cex.lab=1.5,
      cex.axis=1.2, pch=4, col="Gray", xlab=bquote(hat(sigma)[11]),
      ylab=bquote(hat(sigma)[12]), ylim=c(0,5.7))
points(Sigma[1,1], Sigma[1,2], col="red", pch=10)
plot(x=hinnangud_Sigma[,1], y=hinnangud_Sigma[,4], cex.lab=1.5,
      cex.axis=1.2, pch=4, col="Gray", xlab=bquote(hat(sigma)[11]),
      ylab=bquote(hat(sigma)[22]))
points(Sigma[1,1], Sigma[2,2], col="red", pch=10)
plot(x=hinnangud_Sigma[,2], y=hinnangud_Sigma[,4], cex.lab=1.5,
      cex.axis=1.2, pch=4, col="Gray", xlab=bquote(hat(sigma)[12]),
      ylab=bquote(hat(sigma)[22]), xlim=c(0,5))
points(Sigma[1,2], Sigma[2,2], col="red", pch=10)

```

Lisa 2. Kahjude analüüsi R-i kood

```

#setwd("C:/Users/K2trin/Desktop/Loputoo")
#install.packages("openxlsx")
library(openxlsx)
Data=read.xlsx("OCTA.xlsx", sheet=1, rowNames = T, detectDates=T)
head(Data)
names(Data)=c("kahjujuhtum", "valjamakse", "kahjusumma")

as.Date(Data$kahjujuhtum, format="%d/%m/%Y")
as.Date(Data$valjamakse, format="%d/%m/%Y")
library(dplyr)
andmed=arrange(Data, Data$kahjujuhtum)

andmed=mutate(andmed, paevad=as.numeric(valjamakse-kahjujuhtum+1))
filter(andmed, andmed$paevad<=0)
# kahjujuhtum valjamakse kahjusumma paevad
# 1 2001-01-05 1899-12-31      162.31 -36894
# 2 2001-08-17 2001-08-06      297.85  -10
# 3 2002-11-21 2002-11-20       87.42   0

andmed=filter(andmed, andmed$paevad>0)

```

```

summary(andmed)
#kahjujuhtum          valjamakse          kahjusumma          paevad
# Min.      :2001-01-01    Min.      :2001-01-09    Min.      :    0.75    Min.
:      1.0
# 1st Qu.:2001-08-22    1st Qu.:2001-10-30    1st Qu.:   98.00    1st Qu.:
21.0
# Median :2002-02-12    Median :2002-05-03    Median :  183.93    Median :
36.0
# Mean    :2002-02-08    Mean     :2002-04-29    Mean     :  340.29    Mean
:    81.5
# 3rd Qu.:2002-08-19    3rd Qu.:2002-10-31    3rd Qu.:  379.76    3rd Qu.:
78.0
# Max.     :2002-12-31    Max.     :2004-12-31    Max.     :9432.00    Max.
:1142.0

kahjud=select (andmed, kahjujuhtum)%>%
  group_by(kahjujuhtum)%>%
  summarise(sagedus=n())

top_n(kahjud, 3, sagedus)%>%
  arrange(desc(sagedus))
# kahjujuhtum sagedus
# <date>      <int>
# 1 2001-12-19      61
# 2 2001-12-21      39
# 3 2001-12-28      34

library(MASS)
library(fitdistrplus)
#install.packages("actuar")
library(actuar)

##### KAHJUSUMMA – Jaotuste sobitamine empiirilisele jaotusele
x <- seq(0,max(andmed$kahjusumma),length=400)

#Histogrammid
par(mfrow=c(1,2))
hst_kahjusumma <- hist(andmed$kahjusumma, breaks=x, col="blanchedalmond",
  main="", xlab = "Kahjusumma", ylab="Sagedus")
hst_kahjusumma <- hist(andmed$kahjusumma, breaks=x, col="blanchedalmond",
  main="", xlab="Kahjusumma", ylab="Sagedus", xlim=c(0,3000))

```

```

#install.packages("fBasics")
library(fBasics)
basicStats(andmed$kahjusumma) # sd 577, asymmeetriakordaja 7.51

# Sobitame lognormaalse jaotuse andmetele
fitted_ln <- fitdist(andmed$kahjusumma,"lnorm")
# Genereerime v22rtused saadud STM parameetrite hinnangutega
fit_ln <- dlnorm(x, fitted_ln$estimate['meanlog'],
                 fitted_ln$estimate['sdlog'])
fitted_ln #parameetrite hinnangud meanlog=5.236703, sdlog=1.077936

# Sobitame Gammajaotuse andmetele
fitted_gamma <- fitdist(andmed$kahjusumma,"gamma")
# Genereerime v22rtused saadud STM parameetrite hinnangutega
fit_gamma <- dgamma(x, fitted_gamma$estimate['shape'],
                   fitted_gamma$estimate['rate'])
fitted_gamma #parameetrite hinnangud-shape=0.976099289, rate=0.002868064

#Sobitame Weibulli jaotuse andmetele
fitted_weibull <- fitdist(andmed$kahjusumma,"weibull")
# Genereerime v22rtused saadud STM parameetrite hinnangutega
fit_weibull <- dweibull(x, fitted_weibull$estimate['shape'],
                       fitted_weibull$estimate['scale'])
fitted_weibull

#Sobitame Pareto jaotuse andmetele
fitted_pareto <- fitdist(andmed$kahjusumma,"pareto")
# Genereerime v22rtused saadud STM parameetrite hinnangutega
fit_pareto <- dpareto(x, fitted_pareto$estimate['shape'],
                     fitted_pareto$estimate['scale'])
fitted_pareto

par(mfrow=c(1,1))
#Histogramm + jaotusfunktsioonide graafik
hst <- hist(andmed$kahjusumma, breaks=x, col="blanchedalmond", prob=T,
            main="", cex.lab=1.2, xlab = "Kahjusumma", xlim=c(0,3000),
            ylab="Tihedus", cex.axis=1.2)
lines(x, fit_ln, type="l", col="deepskyblue2", lwd=2)
lines(x, fit_gamma, type="l", col="deeppink", lwd=2)
lines(x, fit_weibull, type="l", col="darkorchid1", lwd=2)

```



```

lines(x, fit_pareto, type="l", col="coral1", lwd=2)
legend("topright", legend=c("Gammajaotus", "Weibulli_jaotus", "Pareto_jaotus",
                             "Lognormaalne_jaotus"),
       col=c("deeppink", "darkorchid1", "coral1", "deepskyblue2"), lwd=2,
       cex = 1.2, box.lty = 0)

# Kvantiil-kvantiil graafik
plot.legend=c("Gammajaotus", "Weibulli_jaotus", "Pareto_jaotus",
              "Lognormaalne_jaotus")

qqcomp(main="Kvantiil-kvantiil_graafik", xlab="Sobitatud_jaotuste_kvantiilid",
       ylab="Kahjusement_kvantiilid",
       list(fitted_gamma, fitted_weibull, fitted_pareto, fitted_ln),
       legendtext = plot.legend)

1.39/sqrt(nrow(andmed)) #KS testi kriitiline v22rtus 0.01603
# Kolmogorov-Smirnovi teststatistiku v22rtus ja AIC, BIC
gof_tulemus=gofstat(list(fitted_gamma, fitted_weibull, fitted_pareto, fitted_ln),
                    fitnames = c("gamma", "weibull", "Pareto", "lognorm"))
gof_tulemus
#Kolmogorov-Smirnovi testi tulemused
gof_tulemus$kstest

##### KAHJUMENETLUSE AEG – Jaotuse sobitamise empiirilisele jaotusele
y <- seq(0, max(andmed$paevad), length=200)

par(mfrow=c(1,1))
hst_paevad <- hist(andmed$paevad, breaks=y, col="azure2", ylab="Sagedus",
                  main="", xlab = "Kahjumenetluse_kestus_paevades")
basicStats(andmed$paevad) #sd 123, skew 3.51

# Sobitamise lognormaalse jaotuse andmetele
fitted_ln2 <- fitdist(andmed$paevad, "lnorm")
# Genereerime v22rtused saadud STM parameetrite hinnangutega
fit_ln2 <- dlnorm(y, fitted_ln2$estimate['meanlog'],
                  fitted_ln2$estimate['sdlog'])
fitted_ln2 #parameetrite hinnangud meanlog=5.236703, sdlog=1.077936

# Sobitamise Gammajaotuse andmetele
fitted_gamma2 <- fitdist(andmed$paevad, "gamma")
# Genereerime v22rtused saadud STM parameetrite hinnangutega

```

```

fit_gamma2 <- dgamma(y, fitted_gamma2$estimate['shape'],
                    fitted_gamma2$estimate['rate'])
fitted_gamma2 #parameetrite hinnangud - shape=0.976099289, rate=0.002868064

#Sobitame Weibulli jaotuse andmetele
fitted_weibull2 <- fitdist(andmed$paevad, "weibull")
# Genereerime v22rtused saadud STM parameetrite hinnangutega
fit_weibull2 <- dweibull(y, fitted_weibull2$estimate['shape'],
                        fitted_weibull2$estimate['scale'])

fitted_weibull2

#Sobitame Pareto jaotuse andmetele
fitted_pareto2 <- fitdist(andmed$paevad, "pareto")
# Genereerime v22rtused saadud STM parameetrite hinnangutega
fit_pareto2 <- dpareto(y, fitted_pareto2$estimate['shape'],
                      fitted_pareto2$estimate['scale'])

fitted_pareto2

#Histogramm + jaotusfunktsioonide graafik
hst <- hist(andmed$paevad, breaks=y, col="azure2", prob=T,
            main="", cex.axis=1.2, cex.lab=1.2, xlab = "Kahjumenetluse_aeg",
            ylab="Tihedus")
lines(y, fit_ln2, type="l", col="deepskyblue2", lwd=2)
lines(y, fit_gamma2, type="l", col="deeppink", lwd=2)
lines(y, fit_weibull2, type="l", col="darkorchid1", lwd=2)
lines(y, fit_pareto2, type="l", col="coral1", lwd=2)
legend("topright", legend=c("Gammajaotus", "Weibulli_jaotus", "Pareto_jaotus",
                           "Lognormaalne_jaotus"),
      col=c("deeppink", "darkorchid1", "coral1", "deepskyblue2"), lwd=2,
      cex = 1.2, box.lty = 0)

par(mfrow=c(1,1))
plot.legend=c("Gammajaotus", "Weibulli_jaotus", "Pareto_jaotus",
              "Lognormaalne_jaotus", "inv-gamma")
qqcomp(xlim=c(0,2000), ylim=c(0,1500), cex.lab=1.2, cex.axis=1.2, addlegend=
      main="Kvantiil-kvantiil_graafik", xlab="Sobitatud_jaotuste_kvantiilid",
      ylab="Kahjumenetluse_aja_kvantiilid",
      list(fitted_gamma2, fitted_weibull2, fitted_pareto2, fitted_ln2),
      legendtext = plot.legend)

# Kolmogorov-Smirnovi teststatistiku v22rtus ja AIC, BIC

```

```

gof2=gofstat(list(fitted_gamma2,fitted_weibull2, fitted_pareto2, fitted_ln2)
              fitnames = c("gamma", "weibull", "Pareto", "lognorm"))
gof2
gof2$kstest

##### UHISJAOTUSE LEIDMINE
#library(dplyr)
library(sn)
#esialgsete andmete pealt R, sealt votan tau.
x_tunnus=andmed$kahjusumma #n=7516
y_tunnus=andmed$paevad
yhis=cbind(x_tunnus, y_tunnus)
nu=4 #vabadusastmete arv
p=2 #2-mootmeline juht
#Leiame nihketa hinnangud keskv22rtusele ja kov.maatriksile
X_kaetud=matrix(c(mean(x_tunnus), mean(y_tunnus)), nrow=2, ncol=1)
S=cov(yhis)
S
# x_tunnus y_tunnus
# x_tunnus 333097.826 7140.983
# y_tunnus 7140.983 15299.233

#Kasutame sisendina korrelatsioonimaatriksit R, leiame selle hinnangu
S_d=matrix(c(diag(S)[1],0,0,diag(S)[2]), nrow=2, ncol=2)
R_hinnang=solve(sqrt(S_d))%*%S%*%solve(sqrt(S_d))
R_hinnang
# [,1] [,2]
# [1,] 1.0000000 0.1000316
# [2,] 0.1000316 1.0000000
tau=(2/pi)*asin(R_hinnang[1,2])
tau #0.06378879

# Leiame hinnangud Sigmale ja alphale
nimetaja=as.numeric(1/sqrt(((nu-2)/nu)*
                           (((gamma((nu-1)/2)/gamma(nu/2))^2*(nu-2)/pi)-
                           t(X_kaetud)%*%solve(S+X_kaetud%*%t(X_kaetud))%*%X_kaetud)))
alpha_hinnang=nimetaja*(solve(S+X_kaetud%*%t(X_kaetud))%*%X_kaetud)
alpha_hinnang
# [,1]
# x_tunnus 0.002607586
# y_tunnus 0.013938374

```

```

Sigma_hinnang=((nu-2)/nu)*(S+X_kaetud%*%t(X_kaetud))
Sigma_hinnang
# x_tunnus y_tunnus
# x_tunnus 224447.65 17437.46
# y_tunnus 17437.46 10970.81

##### Asymmeetrilise t-koopula algoritm
St_algoritm=function(R, alpha, nu, p, n){
  #1) Cholesky kompositsioon A Sigmast ( $AA^T=Sigma$ )
  A=t(chol(R))
  X=matrix(c(rep(0, n*2)), nrow=n, ncol=2)
  j=1
  while (j <= n) {
    #2) p solumatut v22rtust  $N(0,1)$ -st vektorisse z
    z=matrix(c(rnorm(p, mean = 0, sd = 1)), nrow=2, ncol=1)
    # 3) Vektor  $x=Az$ 
    x=A%*%z
    # 4)  $z_0 \sim N(0,1)$ 
    z_0=rnorm(1, mean = 0, sd = 1)
    # 5) Moodustame vektori y vastavalt ette antud eeskirjale
    if (t(alpha)%*%x>z_0) {y=x} else {y=-x}
    # 6) Simuleerime h hii-ruut jaotusest
    h=rchisq(1, df=nu)
    # 7) Leiame  $t=y/\sqrt{h/nu}$ 
    t=y/sqrt(h/nu)
    # 8) valime vektori u, nii et  $u_i=G(t_i; 0, \sigma_{ii}, \alpha_i)$ ,
    # $G(*)$  on 1-mootmelise asymmeetrilise t-jaotuse jaotusfunktsioon
    u=matrix(c(rep(0, p)), nrow=p, ncol=1)
    for (i in 1:p){
      u[i]=pst(t[i], omega=R[i], alpha=alpha[i])
    }
    # 9) Moodustame vektori  $x=(F^{-1}_1(u_1), \dots, F^{-1}_p(u_p))$ 
    X_i=matrix(c(rep(0, p)), nrow=1, ncol=p)
    X_i[1]=qlnorm(u[1], fitted_ln$estimate['meanlog'],
                  fitted_ln$estimate['sdlog'])
    X_i[2]=qlnorm(u[2], fitted_ln2$estimate['meanlog'],
                  fitted_ln2$estimate['sdlog'])
    if (X_i[1] != Inf & X_i[2] != Inf) {
      X[j,]=X_i
      j=j+1
    }
  }
}

```

```

    }
  }
  return(X)
}

X_tulemus=St_algoritm(R_hinnang, alpha_hinnang, nu, p, 7500)

par(mfrow=c(1,1))
plot(X_tulemus, main="Simuleeritud_andmed", xlim=c(0,500000), xaxt="none",
      xlab="Kahjusumma", ylab="Kahjumenetluse_aeg", cex.lab=1.2, cex.axis=1.2)
marks=c(seq(0,500000, 100000))
axis(1, at=marks, labels=format(marks, scientific=FALSE), cex.axis=1.2)
plot(X_tulemus)
par(mfrow=c(1,2))
plot(X_tulemus, ylim=c(0,1150), xlim=c(0,9500), main="Simuleeritud_andmed",
      xlab="Kahjusumma_suurus", ylab="Kahjumenetluse_aeg", cex.lab=1.2,
      cex.axis=1.2)
plot(andmed$kahjusumma, andmed$paevad, ylim=c(0,1150), xlim=c(0,9500),
      xlab="Kahjusumma_suurus", ylab="Kahjumenetluse_aeg",
      main="Empiirilised_andmed", cex.lab=1.2, cex.axis=1.2)
cor(X_tulemus, method = "kendall")[1,2] #0.06064065

tau_v22rtused=rep(NA, 1000)
m=1
while (m <= 1000){
  X_tulemus=St_algoritm(R_hinnang, alpha_hinnang, nu, p, 7500)
  tau_v22rtused[m]=cor(X_tulemus, method = "kendall")[1,2]
  m=m+1
}

par(mfrow=c(1,1))
hist(tau_v22rtused, col="blanchedalmond", ylab="Sagedus",
      main=bquote(tau~"vaartused_simuleeritud_valimitel"),
      xlab=bquote(tau~"vaartus"), cex.lab=1.2, cex.axis=1.2)
basicStats(tau_v22rtused)

```

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kättrin Suvi

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Mitmemõõtmeline asümmeetriline $t_{p,\nu}$ -jaotus ja koopula", mille juhendaja on Tõnu Kollo, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kättrin Suvi

15.05.2019